

This Page Is Inserted by IFW Operations  
and is not a part of the Official Record

## BEST AVAILABLE IMAGES

Defective images within this document are accurate representations of the original documents submitted by the applicant.

Defects in the images may include (but are not limited to):

- BLACK BORDERS
- TEXT CUT OFF AT TOP, BOTTOM OR SIDES
- FADED TEXT
- ILLEGIBLE TEXT
- SKEWED/SLANTED IMAGES
- COLORED PHOTOS
- BLACK OR VERY BLACK AND WHITE DARK PHOTOS
- GRAY SCALE DOCUMENTS

## IMAGES ARE BEST AVAILABLE COPY.

As rescanning documents *will not* correct images,  
Please do not report the images to the  
Image Problem Mailbox.

**THIS PAGE BLANK (USPTO)**

PCT

WORLD INTELLECTUAL PROPERTY ORGANIZATION  
International Bureau



INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) International Patent Classification 6: <b>C12N 15/57, 9/54, 1/21, C12P 21/06</b>		A1	(11) International Publication Number: <b>WO 96/27671</b>
			(43) International Publication Date: 12 September 1996 (12.09.96)
(21) International Application Number: <b>PCT/US96/02861</b>		(81) Designated States: AU, CA, JP, European patent (AT, BE, CH, DE, DK, ES, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE).	
(22) International Filing Date: 27 February 1996 (27.02.96)		Published With international search report. Before the expiration of the time limit for amending the claims and to be republished in the event of the receipt of amendments.	
(30) Priority Data: 08/398,028 3 March 1995 (03.03.95) US 08/504,265 19 July 1995 (19.07.95) US			
(71) Applicant: GENENTECH, INC. [US/US]; 460 Point San Bruno Boulevard, South San Francisco, CA 94080-4990 (US).			
(72) Inventors: BALLINGER, Marcus, D.; 918 Chula Vista Avenue #3, Burlingame, CA 94010 (US). WELLS, James, A.; 1342 Columbus Avenue, Burlingame, CA 94010 (US).			
(74) Agents: KUBINEC, Jeffrey, S. et al.; Genentech, Inc., 460 Point San Bruno Boulevard, South San Francisco, CA 94080-4990 (US).			

(54) Title: SUBTILISIN VARIANTS CAPABLE OF CLEAVING SUBSTRATES CONTAINING BASIC RESIDUES

(57) Abstract

The bacterial serine protease, subtilisin BPN', has been mutated so that it will efficiently and selectively cleave substrates containing basic residues. Combination mutants, where Asn 62 was changed to Asp, Gly 166 was changed to Asp (N62D/G166D), and optionally Tyr 104 was changed to Asp had a larger than additive shift in specificity toward substrates containing basic residues. Suitable substrates of the subtilisin variants were revealed by sorting a library of phage particles (substrate phage) containing five contiguous randomized residues. This method identified a particularly good substrate, Asn-Leu-Met-Arg-Lys- (SEQ ID NO: 35), that was selectively cleaved in the context of a fusion protein by the N62D/G166D subtilisin variant. A particularly good substrate for N62D/G166D/Y104D would be Asn-Arg-Met-Arg-Lys- (SEQ ID NO: 76). Accordingly, these subtilisin variants are useful for cleaving fusion proteins with basic substrate linkers and processing hormones or other proteins *in vitro* or *in vivo* that contain basic cleavage sites.

**FOR THE PURPOSES OF INFORMATION ONLY**

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

AM	Armenia	GB	United Kingdom	MW	Malawi
AT	Austria	GZ	Georgia	MX	Mexico
AU	Australia	GN	Guinea	NZ	Niger
BB	Barbados	GR	Greece	NL	Netherlands
BE	Belgium	HU	Hungary	NO	Norway
BF	Burkina Faso	IU	Ireland	NZ	New Zealand
BG	Bulgaria	IT	Italy	PL	Poland
BJ	Benin	JP	Japan	PT	Portugal
BR	Brazil	KZ	Kenya	RO	Romania
BY	Belarus	KG	Kyrgyzstan	RU	Russian Federation
CA	Canada	KP	Democratic People's Republic of Korea	SD	Sudan
CF	Central African Republic	KR	Republic of Korea	SE	Sweden
CG	Congo	KZ	Kazakhstan	SG	Singapore
CH	Switzerland	L1	Liechtenstein	SI	Slovenia
CI	Côte d'Ivoire	LX	Sri Lanka	SK	Slovakia
CM	Cameroon	LR	Liberia	SN	Senegal
CN	China	LT	Lithuania	SZ	Senegal
CS	Czechoslovakia	LU	Luxembourg	TD	Chad
CZ	Czech Republic	LV	Latvia	TG	Togo
DE	Germany	MC	Monaco	TJ	Tajikistan
DK	Denmark	MD	Republic of Moldova	TT	Trinidad and Tobago
EE	Estonia	MG	Madagascar	UA	Ukraine
ES	Spain	ML	Mali	UG	Uganda
FI	Finnland	MN	Mongolia	US	United States of America
FR	France	MR	Mauritania	UZ	Uzbekistan
GA	Gabon			VN	Viet Nam

## INTERNATIONAL SEARCH REPORT

Information on patent family members

International Application No  
PCT/US-96/02861

Patent document cited in search report	Publication date	Patent family member(s)		Publication date
EP-A-0316748	24-05-89	DE-A- 3739347 AU-B- 2569588 CA-A- 1337283 DE-A- 3873273 ES-T- 2051816 IE-B- 62228 JP-A- 1160496 US-A- 5270176		01-06-89 29-06-89 10-10-95 03-09-92 01-07-94 11-01-95 23-06-89 14-12-93
WO-A-9111454	08-08-91	AU-B- 7166891 DE-D- 69101486 DE-T- 69101486 EP-A- 0511978 ES-T- 2062759 JP-T- 5503631		21-08-91 28-04-94 04-08-94 11-11-92 16-12-94 17-06-93
WO-A-9530010	09-11-95	AU-B- 2159195 AU-B- 2293195 WO-A- 9529979 AU-B- 7870394 CA-A- 2170491 EP-A- 0719339 NO-A- 961067 WO-A- 9507991		29-11-95 29-11-95 09-11-95 03-04-95 23-03-95 03-07-96 15-05-96 23-03-95
EP-A-0405901	02-01-91	CA-A- 2034486 WO-A- 9100334 JP-T- 4500385 EP-A- 0405902 WO-A- 9100335 JP-T- 4500384		27-12-90 10-01-91 23-01-92 02-01-91 10-01-91 23-01-92

**SUBTILISIN VARIANTS CAPABLE OF CLEAVING SUBSTRATES  
CONTAINING BASIC RESIDUES**

**FIELD OF THE INVENTION**

This invention relates to subtilisin varianis having altered specificity from wild-type subtilisins.

5 Specifically, the subtilisin variants are modified so that they efficiently and selectively cleave substrates containing basic residues. The invention further relates to the DNA encoding these novel polypeptides, as well as the recombinant materials and methods for producing these subtilisin variants. In a particular aspect, the present invention provides for processes for cleaving protein substrates containing basic residues.

**BACKGROUND OF THE INVENTION**

10 Site-specific proteolysis is one of the most common forms of post-translational modifications of proteins (for review see Neurath, H. (1989) Trends Biochem. Sci., 14:268). In addition, proteolysis of fusion proteins *in vitro* is an important research and commercial tool (for reviews see Uhlen, M. and Moks, T. (1990) Methods Enzymol., 185:129-143; Carter, P. (1990) in Protein Purification: From Molecular Mechanisms to Large-Scale Processes, M.R. Landisch, R.C. Wilson, C.D. Painton, S.E. Builder, Eds. (ACS Symposium Series 427, 15 American Chemical Society, Washington, D.C.), Chap. 13, p.181-193; and Nilsson, B. et al. (1992) Current Opin. Struct. Biol., 2:569). Expressing a protein of interest as a fusion protein facilitates purification when the fusion contains an affinity domain such as glutathione-S-transferase, Protein A or a poly-histidine tail. The fusion domain can also facilitate high level expression and/or secretion.

To liberate the protein product from the fusion domain requires selective and efficient cleavage of the fusion protein. Both chemical and enzymatic methods have been proposed (see references above). Enzymatic methods are generally preferred as they tend to be more specific and can be performed under mild conditions that avoid denaturation or unwanted chemical side-reactions. A number of natural and even designed enzymes have been applied for site-specific proteolysis. Although some are generally more useful than others (Forsberg, G., Bastrup, B., Rondahl, H., Holmgren, E., Pohl, G., Hartmanis, M. and Lake, M. (1992) J. Prot. Chem., 25 11:201-211), no one is applicable to every situation given the sequence requirements of the fusion protein junction and the possible existence of protease sequences within the desired protein product. Thus, an expanded array of sequence specific proteases, analogous to restriction endonucleases, would make site-specific proteolysis a more widely used method for processing fusion proteins or generating protein/peptide fragments either *in vitro* or *in vivo*.

30 The processing of prohormones by the KEX2-related family of serine endoproteases illustrates one of the most precise proteolytic events found in nature (for reviews see Steiner, D. F., Smeeekens, S. P., Ohagi, S. and Chan, S. J. (1992) J. Biol. Chem., 267, 23435-23438 and Smeeekens, S. P. (1993) Bio/Technology 11, 182-186). This family of proteases, that includes the yeast KEX2 and the mammalian PC2, PC3 and furin enzymes, are homologous to the bacterial serine protease subtilisin (Kraut, J. (1977) Annu. Rev. Biochem., 46:331-358).  
35 Subtilisin has a broad substrate specificity that reflects its role as a scavenger protease. In contrast, these eukaryotic enzymes are very specific for cleaving substrates containing two basic residues and thus well-suited for site-specific proteolysis.

All of these eucaryotic enzymes strongly require Arg at the P1 position, and either Arg, Lys or Pro at the P2 position of peptide substrates. The prohormone convertases from higher eukaryotes such as furin, PC2, and PC3 also have an absolute requirement for Arg at the P4 position (Bresnahan, P. A., Leduc, R., Thomas, L., Thorner, J., Gibson, H. L., Brake, A. J., Barr, P. J. and Thomas, G. (1990) *J. Cell. Biol.* 111, 2851; Wise, R. J.,

5 Baar, P. J., Wong, P. A., Kiefer, M. C., Brake, A. J., and Kaufman, R. J. (1990) *Proc. Natl. Acad. Sci. USA* 87, 9378-9382.; Hosaka, M., Nagahama, M., Kim, W.-S., Watanabe, T., Hatsuzakawa, K., Ikemizu, J., Murakami, K., and Nakayama, K. (1991) *J. Biol. Chem.* 266, 12127-12130.; Matthews, D. J., Goodman, L. J., Gorman, C. M., and Wells, J. A. (1994) *Protein Science* 3, 1197-1205).

Despite the very narrow specificity of the pro-hormone processing enzymes, in some cases they are  
10 capable of rapid cleavage of target sequences. For example, the  $k_{cat}/K_m$  ratio for KEX2 to cleave a good substrate (e.g. acetyl-pMYRK-MCA) is  $1.1 \times 10^7 \text{ M}^{-1}\text{s}^{-1}$  (Brenner, C., and Fuller, R.S. (1992) *Proc. Natl. Acad. Sci. USA*, 89:922-926) compared to  $3 \times 10^3$  for subtilisin cleaving a good substrate (e.g. suc-AAPF-pNA) (Estell, D. A., Graycar, T. P., Miller, J. V., Powers, D. B., Burnier, J. P., Ng, P. G. and Wells, J.A. (1986) *Science*, 233:659-663).

15 However, the eukaryotic proteases are expressed in small amounts (Bravo, D. B., Gleason, J. B., Sanchez, R. I., Roth, R. A., and Fuller, R. S. (1994) *J. Biol. Chem.*, 269:25830-25837 and Matthews, D. J., Goodman, L. J., Gorman, C. M., and Wells, J. A. (1994) *Protein Science* , 3:1197-1205) making them impractical to apply presently to processing of fusion proteins *in vitro*. Subtilisin BPN' however, can be expressed in large amounts (Wells, J.A., Ferrari, E., Henner, D.J., Estell, D.A. and Chen, E.Y. (1983) *Nucl. Acids Res.*, 11:7911-7929)

Extensive protein engineering studies of subtilisin, and especially subtilisin BPN', have identified several residues in the S1 and S2 active site of the enzyme where amino acid substitutions lead to large changes in substrate specificity (Wells, J. A., and Estell, D.A., (1988) *Trends Biochem. Sci.*, 13:291-297; Carter, P., et al., (1989) *PROTEINS:Structure, Function, and Genetics*, 6:240-248). X-ray crystal structures of subtilisin  
25 containing bound transition state analogues (Wright, C. S., Alden, R. A. and Kraut, J. (1969) *Nature* , 221:235-242; McPhalen, C.A. and James, N.G. (1988) *Biochemistry*, 27:6582-6598; Bode, W., Papamokos, E., Musil, D., Seemueller, U. and Fritz, M. (1986) *EMBO J.*, 5:813-818; and Bott, R., Uhtsch, M., Kossiakoff, A., Graycar, T., Katz, B. and Power, S. (1988) *J. Biol. Chem.*, 263:7895-7906) can be used to locate active site residues that are in close proximity to side chains at key positions in substrate peptides (Wells, J.A., (1987) *Proc. Natl. Acad. Sci. USA* 84:1219-1223). Consideration of electrostatic interactions between charged peptide substrates and subtilisin can be used to tailor the substrate binding cleft of the subtilisin BPN' to favor complementary charged substrates (Wells, J.A., et al., (1987) *Proc. Natl. Acad. Sci., USA*, 84:1219-1223). Previous work has shown that replacement of residues at position 156 and 166 in the S1 binding site of subtilisin BPN' with various charged residues leads to improved specificity for complementary charged substrates.

35 A substantial amount of protein engineering has been applied to the specificity determinants of the S4 subsite of subtilisin BPN' in efforts to alter specificity for P4 substrates (Eder, J., Rheinheimer, M., and Fersht, A. R. (1993) *FEBS Lett* 335, 349-352; Rheinheimer, M., Baker, G., Eder, J., and Fersht, A. R. (1993) *Biochemistry* 32, 1199-1203; Rheinheimer, M., Eder, J., Pandey, P.S., and Fersht, A. R. (1994) *Biochemistry* 33,

221-225). However, the mutations introduced consisted entirely of hydrophobic substitutions, thus preserving the overall hydrophobic substrate preference in the site.

Previous attempts to introduce, remove or reverse charge specificity in enzyme active sites have been met with considerable difficulty. This has generally been attributed to a lack of stabilization of the introduced charge or enzyme-substrate ion pair complex by the wild-type enzyme environment (Hwang, J.K. and Warshel, A. (1988) *Nature*, 334:270-272). For example, Stennicke *et. al* (Stennicke, H.R.; Ujie, H.M.; Christensen, U.; Remington, S.J.; and Breddam (1994) *Prot. Eng.* 7:911-916) made acidic (D/E) mutations at five residues in the P1' binding of carboxypeptidase Y in an attempt to change the P1' preference from Phe to Lys/Arg. Only the L272D and L272E mutations were found to alter the specificity in the desired direction, up to 1.5-fold preference in Lys/Arg over Phe, and the others simply resulted in less active enzymes having substrate preferences similar to wild-type. In the case of trypsin, a protease that is highly specific for basic P1 residues, recruitment of chymotrypsin-like (hydrophobic P1) specificity required not only mutations of the ion pair-forming Asp 189 to Ser, but also transplantation of two more distant surface loops from chymotrypsin (Graf, L., Jancso, A., Szilagyi, L., Hegyi, G., Pinter, K., Naray-Szabo, G., Hepp, J., Medzihradsky, K., and Rutter, W. J., *Proc. Natl. Acad. Sci. USA* (1988) 85:4961-4965 and Hedstrom, L., Szilagyi, L., and Rutter, W. J., *Science* (1992) 255:1249-1253).

In the present work, we have also verified that relatively low specificity is gained by introducing single ion-pairs between enzyme and substrate. However, when two or more choice ionic interactions were simultaneously engineered into subtilisin BPN', the resulting variants had higher specificity for basic residues in each of the subsites due to a non additive effect.

Accordingly, it is an object to produce a subtilisin variant with basic specificity for use in processing pro-proteins made by recombinant techniques.

#### SUMMARY OF THE INVENTION

The present invention provides for subtilisin variants with altered substrate specificity. Preferred subtilisin variants are highly specific for the efficient cleavage of substrates containing basic residues. The subtilisin variants have a substrate specificity which is substantially different from the substrate specificity of the precursor subtilisin from which the amino acid sequence of the variant is derived. The amino acid sequence of the subtilisin variants are derived by the substitution of one or more amino acids of a precursor subtilisin amino acid sequence.

In a preferred aspect of the present invention, the subtilisin variants of the present invention are specific for the cleavage of protein substrates containing basic amino acid residues at substrate positions P1, P2 and P4. According to this aspect of the present invention subtilisin variants having amino acid substitutions at positions corresponding to amino acid positions 62, 104 and 166 of subtilisin BPN' produced by *Bacillus amyloliquefaciens* are preferred. Accordingly, subtilisin variants are provided wherein amino acids 62, 104 and 166 of subtilisin BPN' are substituted with an acidic amino acids. Preferably the acidic amino acid is Asp or Glu, and most preferably Asp.

WO 96/27671

Preferred substrates for the subtilisin variants according to this aspect of the present invention contain either Lys (K) or Arg (R) at substrate positions P2 and P1, practically any residue at P3, and preferably either Lys or Arg at P4, and again practically any residue at P5. Thus an exemplary good substrate would contain -Asn-Arg-Met-Arg-Lys- (SEQ ID NO: 76) at -P5-P4-P3-P2-P1- respectively. Additionally, good substrates would not have Pro at P1', P2', or P3' nor would Ile be present at P1'.

According to a second aspect of the present invention the subtilisin variants are capable of cleaving protein substrates having basic residues at positions P1 and P2. According to this aspect of the present invention subtilisin variants having amino acid substitutions at positions corresponding to amino acid positions 62, and 166 of subtilisin BPN' produced by *Bacillus amyloliquefaciens* are preferred. The preferred subtilisin variants having substrate specificity for dibasic substrates have an acidic amino acid residue at residue position 62 of subtilisin naturally produced by *Bacillus amyloliquefaciens*. In a preferred embodiment, the naturally occurring Asn at residue position 62 of subtilisin BPN' is preferably substituted with an acidic amino acid residue such as Glu or Asp, and most preferably Asp. The preferred subtilisin variants, having substrate specificity for substrates having dibasic amino acid residues, additionally have an acidic residue, Asp or Glu, at residue position 166 of subtilisin BPN'. Thus, the subtilisin BPN' variant containing substitution of amino acids 62 and 166 with acidic amino acids Glu or Asp are preferred. In particular, a subtilisin variant having amino acid Asp at positions 62 and 166 is preferred (subtilisin BPN' variant N62D/G166D). The subtilisin variants according to this aspect of the invention may be used to cleave substrates containing dibasic residues such as fusion proteins with dibasic substrate linkers and processing hormones or other proteins (*in vitro* or *in vivo*) that contain dibasic cleavage sites.

Preferred substrates for the subtilisin BPN' variant N62D/G166D contain either Lys (K) or Arg (R) at substrate positions P2 and P1, practically any residue at P3, a non-charged hydrophobic residue at P4, and again practically any residue at P5. Thus an exemplary good substrate would contain -Asn-Leu-Met-Arg-Lys-(SEQ ID NO: 35) at -P5-P4-P3-P2-P1- respectively. Additionally, good substrates would not have Pro at P1', P2', or P3' nor would Ile be present at P1'.

The invention also includes mutant DNA sequences encoding such subtilisin variants. These mutant DNA sequences are derived from a precursor DNA sequence which encodes a naturally occurring or recombinant precursor subtilisin. The mutant DNA sequence is derived by modifying the precursor DNA sequence to encode the substitution(s) of one or more amino acids encoded by the precursor DNA sequence. These recombinant DNA sequences encode mutants having an amino acid sequence which does not exist in nature and a substrate specificity which is substantially different from the substrate specificity of the precursor subtilisin encoded by the precursor DNA sequence.

Further the invention includes expression vectors containing such mutant DNA sequences as well as host cells transformed with such vectors which are capable of expressing the subtilisin variants.

The invention also provides for a process for cleaving a polypeptide such as a fusion protein containing a substrate linker represented by the formula:

P4-P3-P2-P1

wherein P4 is a basic amino acid or a large hydrophobic amino acid such as Leu or Met; P3 is an amino acid selected from the naturally occurring amino acids; P2 is a basic amino acid; and P1 is a basic amino acid. The

process includes the step of subjecting the polypeptide to the subtilisin variants described herein under conditions such that the subtilisin variant cleaves the polypeptide.

#### BRIEF DESCRIPTION OF THE FIGURES

Figure 1. Structure of a succinyl-Ala-Ala-Pro-BoroPhe (SEQ ID NO: 69) inhibitor bound to the active site of subtilisin BPN' showing the S2 and S1 binding pocket residues subjected to mutagenesis.

Figure 2. Kinetic analysis of S1 binding site subtilisin mutants versus substrates having variable P1 residues. The kinetic constant  $k_{cat}/Km$  was determined from plots of initial rates versus substrate concentration for the tetrapeptide series succinyl-Ala-Ala-Pro-Xaa-pNa (SEQ ID NO: 69), where Xaa was Lys (SEQ ID NO: 58), Arg (SEQ ID NO: 59), Phe (SEQ ID NO: 56), Met (SEQ ID NO: 60) or Gln (SEQ ID NO: 61) (defined to 10 the right of the plot).

Figure 3. Kinetic analysis of S2 binding site subtilisin mutants versus substrates having variable P2 residues. The kinetic constant  $k_{cat}/Km$  was determined from plots of initial rates versus substrate concentration for the tetrapeptide series succinyl-Ala-Ala-Xaa-Phe-pNa (SEQ ID NO: 70), where Xaa was Lys (SEQ ID NO: 62), Arg (SEQ ID NO: 64), Ala (SEQ ID NO: 63), Pro (SEQ ID NO: 56), or Asp (SEQ ID NO: 65) (defined on 15 the right of the plot).

Figure 4. Kinetic analysis of combined S1 and S2 binding site subtilisin mutants versus substrates having variable P1 and P2 residues. The kinetic constants  $k_{cat}/Km$  were determined from plots of initial rates versus substrate concentration for the tetrapeptide series succinyl-Ala-Ala-Xaa<sub>1</sub>-Xaa<sub>2</sub>-pNa (SEQ ID NO: 71), where Xaa<sub>2</sub>-Xaa<sub>1</sub> was Lys-Lys (SEQ ID NO: 66), Lys-Arg (SEQ ID NO: 67), Lys-Phe (SEQ ID NO: 62), Pro-Lys (SEQ 20 ID NO: 58), Pro-Phe (SEQ ID NO: 56), or Ala-Phe (SEQ ID NO: 63) (defined on the right of the plot).

Figure 5. Results of hGH-AP fusion protein assay. hGH-AP fusion proteins were constructed, bound to hGHbp-coupled resin, and treated with 0.5 nM N62D/G166D subtilisin in 20 mM Tris-Cl pH 8.2. Aliquots were withdrawn at various times and AP release was monitored by activity assay in comparison to a standard curve. Arrows indicate the cleavage site. The rate of cleavage of fusion proteins containing various substrate linkers is 25 shown. Substrates containing a Pro at position P1' are not cleaved.

Figure 6-1 - 6-10. (Collectively referred to herein as Fig. 6). DNA sequence of the phagemid pSS5 containing the N62D/G166D double mutant subtilisin BPN' gene (SEQ ID NO: 1), and translated amino acid sequence for the mutant preprosubtilisin (SEQ ID NO: 2). The pre region is comprised of residues -107 to -78, the pro of residues -77 to -1, and the mature enzyme of residues +1 to +275 (SEQ ID NO: 72). Also shown are 30 restriction sites recognized by endonucleases that require 6 or more specific bases in succession.

Figure 7. Structure of a succinyl-Ala-Ala-Pro-BoroPhe (SEQ ID NO: 69) inhibitor bound to the active site of subtilisin BPN' showing the S1, S2, and S4 binding pocket residues subjected to mutagenesis.

Figure 8. DNA sequence of the N62D/Y104D/G166D triple mutant (SEQ ID NO: 74) as well as the 35 translated amino acid sequence (SEQ ID NO: 75). The preregion is comprised of residues -107 to -78, the pro residues -77 to -1 and the mature enzyme +1 to +275. The preregion reflects the changes, A(-4)R/A(-2)K/Y(-1)R made in the wild-type processing site to affect expression.

## DETAILED DESCRIPTION OF THE INVENTION

*Definitions*

Terms used in the claims and specification are defined as set forth below unless otherwise specified.

The term amino acid or amino acid residue, as used herein, refers to naturally occurring L amino acids or residues, unless otherwise specifically indicated. The commonly used one- and three-letter abbreviations for amino acids are used herein (Lehninger, A. L., *Biochemistry*, 2d ed., pp. 71-92, Worth Publishers, N. Y. (1975)). Basic amino acids are Arg and Lys. Acidic amino acids are Asp and Glu.

Substrates are described in triplet or single letter code as Pn...P2-P1-P1'-P2'...Pn'. The "P1" residue refers to the position proceeding (i.e., N-terminal to) the scissile peptide bond (i.e. between the P1 and P1' residues) of the substrate as defined by Schechter and Berger (Schechter, I. and Berger, A., *Biochem. Biophys. Res. Commun.* 27: 157-162 (1967)). Similarly, the term P1' is used to refer to the position following (i.e., C-terminal to) the scissile peptide bond of the substrate. Increasing numbers refer to the next consecutive position preceding (e.g., P2 and P3) and following (e.g., P2' and P3') the scissile bond. According to the present invention the scissile peptide bond is that bond that is cleaved by the subtilisin variants of the instant invention.

"Subtilisins," "precursor subtilisin" and the like are bacterial carbonyl hydrolases which generally act to cleave peptide bonds of proteins or peptides. As used herein, "subtilisin" means a naturally occurring subtilisin or a recombinant subtilisin. A series of naturally occurring subtilisins are known to be produced and often secreted by various bacterial species (Siezen, R.J., et al., (1991) *Protein Engineering* 4:719-737). Amino acid sequences of the members of this series are not entirely homologous. However, the subtilisins in this series exhibit the same or similar type of proteolytic activity. This class of serine proteases shares a common amino acid sequence defining a catalytic triad which distinguishes them from the chymotrypsin related class of serine proteases. The subtilisins and chymotrypsin related serine proteases both have a catalytic triad comprising aspartate, histidine and serine. In the subtilisin related proteases the relative order of these amino acids, reading from the amino to carboxy terminus is aspartate-histidine-serine. In the chymotrypsin related proteases the relative order, however is histidine-aspartate-serine. Thus, subtilisins as used herein refer to a serine protease having the catalytic triad of subtilisin related proteases.

Generally, subtilisins are serine endoproteases having molecular weights of about 27,500 which are secreted in large amounts from a wide variety of *Bacillus* species. The protein sequence of subtilisins have been determined from at least four different species of *Bacillus* (Markland, F.S., et al. (1971) in *The Enzymes*, ed. Boyer P.D., Acad Press, New York, Vol. III, pp. 561-608; and Nedkov, P. et al. (1983) *Hoppe-Seyler's Z. Physiol. Chem.* 364:1537-1540). The three-dimensional crystallographic structure of four subtilisins have been reported (BPN' from *Bacillus amyloliquefaciens*, Hirono et al. (1984) *J. Mol. Biol.* 178:389-413; subtilisin Carlesberg from *Bacillus licheniformis*, Bode et al., (1986) *EMBO J.*, 5:813-818; thermitase from *Thermoactinomyces vulgaris*, Gros et al., (1989) *J. Mol. Biol.* 210:347-367; and proteinase K from *Tritirachium album*, Betzel, et al., (1988) *Acta Crystallogr.*, B, 44:163-172). The three dimensional structure of subtilisin BPN' (from *B. amyloliquefaciens*) to 2.5 Å resolution has also been reported by Wright, C.S. et al. (1969) *Nature* 221:235-242 and Drent, J. et al. (1972) *Eur. J. Biochem.* 26:177-181. These studies indicate that although subtilisin is genetically unrelated to the mammalian serine proteases, it has a similar fold and active site structure. The x-ray crystal structures of subtilisin containing covalently bound peptide inhibitors (Robertus, J.D., et al.

(1972) *Biochemistry* 11:2439-2449), product complexes (Roberts, J.D., et al. (1972) *Biochemistry* 11:4293-4303), and transition state analogs (Matthews, D.A., et al. (1975) *J. Biol. Chem.* 250:7120-7126 and Poulos, T.L., et al. (1976) *J. Biol. Chem.* 251:1097-1103), which have been reported have also provided information regarding the active site and putative substrate binding cleft of subtilisins. In addition, a large number of kinetic and chemical modification studies have been reported for subtilisins (Phillip, M., et al. (1983) *Mol. Cell. Biochem.* 51:5-32; Svendsen, I.B. (1976) *Carlsberg Res. Comm.* 41:237-291 and Markland, F.S. *Id.*) as well as at least one report wherein the side chain of methionine at residue 222 of subtilisin was converted by hydrogen peroxide to methionine-sulfoxide (Stauffer, D.C., et al. (1965) *J. Biol. Chem.* 244:5333-5338).

"Subtilisin variant," "subtilisin mutant" and the like refer to a subtilisin-type serine protease having a sequence which is not found in nature that is derived from a precursor subtilisin according to the present invention. The subtilisin variant has a substrate specificity different from the precursor subtilisin by virtue of amino acid substitutions within the precursor subtilisin amino acid sequence. The term is meant to include subtilisin variants in which the DNA sequence encoding the precursor subtilisin is modified to produce a mutant DNA sequence which encodes the substitution of one or more amino acids in the naturally occurring subtilisin amino acid sequence. Suitable methods to produce such modification include those disclosed in U. S. Patent No. 4,760,025 and 5,371,008 and in EPO Publication No. 0130756 and 0251446.

A change in substrate specificity is defined as a difference between the  $K_m/K_m$  ratio of the precursor subtilisin and the subtilisin variant. The  $K_m/K_m$  ratio is a measure of catalytic efficiency. Subtilisin variants with increased or decreased  $K_m/K_m$  ratios compared to the precursor subtilisin from which they were derived are described herein. Generally, the objective is to secure a variant having a greater, i.e. numerically larger,  $K_m/K_m$  ratio for a given substrate. A greater  $K_m/K_m$  ratio for a particular substrate indicates that the variant may be used to more efficiently cleave the target substrate.

The specificity or discrimination between two or more competing substrates is determined by the ratios of  $k_{cat}/K_m$  (Fersht, A.R., (1985) in *Enzyme Structure and Mechanism*, W.F. Freeman and Co., N.Y. p. 112). An increase in  $K_m/K_m$  ratio for one substrate may be accompanied by a reduction in  $K_m/K_m$  ratio for another substrate. This shift in substrate specificity indicates that the variant subtilisin with the increased  $K_m/K_m$  ratio for the substrate has utility in cleaving the particular substrate over the precursor subtilisin in, for example, preventing undesirable hydrolysis of a particular substrate in a mixture of substrates.

In general, for a subtilisin variant to have a useful catalytic efficiency for cleavage of a particular substrate the  $K_m/K_m$  ratio will generally be between  $1 \times 10^3 \text{ M}^{-1}\text{s}^{-1}$  to about  $1 \times 10^7 \text{ M}^{-1}\text{s}^{-1}$ . More often, the  $K_m/K_m$  ratio will be between about  $1 \times 10^4 \text{ M}^{-1}\text{s}^{-1}$  and  $1 \times 10^6 \text{ M}^{-1}\text{s}^{-1}$ .

When referring to mutants or variants, the wild type amino acid residue is followed by the residue number and the new or substituted amino acid residue. For example, substitution of D for wild type N in residue position 62 is denominated N62D.

"Subtilisin variants or mutants" are designated in the same manner by using the single letter amino acid code for the wild-type residue followed by its position and the single letter amino acid code of the replacement residue. Multiple mutants are indicated by component single mutants separated by slashes. Thus the subtilisin BPN' variant N62D/G166D is a di-substituted variant in which Asp replaces Asn and Gly at residue positions 62 and 166, respectively, in wild-type subtilisin BPN'.

WO 96/27671

An amino acid residue of a precursor carbonyl hydrolase is "equivalent" to a residue of *B. amyloliquefaciens* subtilisin if it is either homologous (i.e., corresponding in position in either primary or tertiary structure) or analogous to a specific residue or portion of that residue in *B. amyloliquefaciens* subtilisin (i.e., having the same or similar functional capacity to combine, react, or interact chemically).

5 In order to establish homology to primary structure, the amino acid sequence of a precursor carbonyl hydrolase is directly compared to the *B. amyloliquefaciens* subtilisin primary sequence and particularly to a set of residues known to be invariant in all subtilisins for which the sequences are known (see e.g. Figure 5-C in EPO 0251446). After aligning the conserved residues, allowing for necessary insertions and deletions in order to maintain alignment (i.e., avoiding the elimination of conserved residues through arbitrary deletion and insertion),  
 10 the residues equivalent to particular amino acids in the primary sequence of *B. amyloliquefaciens* subtilisin are defined. Alignment of conserved residues should conserve 100% of such residues. However, alignment of greater than 75% or as little as 50% of conserved residues is also adequate to define equivalent residues. Conservation of the catalytic triad, Asp32/His64/Ser221, is required.

Equivalent residues homologous at the level of tertiary structure for a precursor carbonyl hydrolase whose  
 15 tertiary structure has been determined by x-ray crystallography, are defined as those for which the atomic coordinates of 2 or more of the main chain atoms of a particular amino acid residue of the precursor carbonyl hydrolase and *B. amyloliquefaciens* subtilisin (N on N, CA on CA, C on C, and O on O) are within 0.13nm and preferably 0.1nm after alignment. Alignment is achieved after the best model has been oriented and positioned to give the maximum overlap of atomic coordinates of non-hydrogen protein atoms of the carbonyl hydrolase  
 20 in question to the *B. amyloliquefaciens* subtilisin. The best model is the crystallographic model giving the lowest R factor for experimental diffraction data at the highest resolution available.

$$R \text{ factor} = \frac{\sum_{\mathbf{h}} |F_o(\mathbf{h})| - |F_c(\mathbf{h})|}{\sum_{\mathbf{h}} |F_o(\mathbf{h})|}$$

Equivalent amino acid residues of subtilisin BPN', subtilisin Carlsberg, thermitase and proteinase K from tertiary structure analysis is provided in, for example, Siezen, et al., (1991) Prot. Eng. 4:719-737.

30 Equivalent residues which are functionally analogous to a specific residue of *B. amyloliquefaciens* subtilisin are defined as those amino acids of the precursor carbonyl hydrolases which may adopt a conformation such that they either alter, modify or contribute to protein structure, substrate binding or catalysis in a manner defined and attributed to a specific residue of the *B. amyloliquefaciens* subtilisin as described herein. Further, they are those residues of the precursor carbonyl hydrolase (for which a tertiary structure has been obtained by  
 35 x-ray crystallography), which occupy an analogous position to the extent that although the main chain atoms of the given residue may not satisfy the criteria of equivalence on the basis of occupying a homologous position, the atomic coordinates of at least two of the side chain atoms of the residue lie within 0.13nm of the

corresponding side chain atoms of *B. amyloliquefaciens* subtilisin. The three dimensional structures would be aligned as outlined above.

Some of the residues identified for substitution are conserved residues whereas others are not. In the case of residues which are not conserved, the replacement of one or more amino acids is limited to substitutions which

5 produce a mutant which has an amino acid sequence that does not correspond to one found in nature. In the case of conserved residues, such replacements should not result in a naturally occurring sequence. The subtilisin mutants of the present invention include the mature forms of subtilisin mutants as well as the pro- and prepro-forms of such subtilisin mutants. The prepro-forms are the preferred construction since this facilitates the expression, secretion and maturation of the subtilisin mutants.

10 "Prosequence" refers to a sequence of amino acids bound to the N-terminal portion of the mature form of a subtilisin which when removed results in the appearance of the "mature" form of the subtilisin. Many proteolytic enzymes are found in nature as translational proenzyme products and, in the absence of post-translational processing, are expressed in this fashion. The preferred prosequence for producing subtilisin mutants, specifically subtilisin BPN' mutants, is the putative prosequence of *B. amyloliquefaciens* subtilisin  
15 although other subtilisin prosequences may be used. For example, when the substrate specificity of the precursor subtilisin is altered according to the present invention, this alteration may affect the ability of the variant subtilisin to undergo autolytic cleavage of the naturally occurring prosequence. In order to affect the expression and proper folding of a mature variant subtilisin whose substrate specificity has been altered, it may be necessary to alter the prosequence to correspond to the new or variant substrate specificity.

20 As an example, the substrate specificity of a particular subtilisin variant N62D/Y104D/G166D is distinct from the precursor subtilisin from which it was derived. The subtilisin variant prefers substrates containing basic residues at substrate positions corresponding to P4, P2, and P1. According to this aspect of the present invention, the precursor prosequence which was efficiently autolyzed by the precursor subtilisin is altered to correspond to the substrate specificity of the variant subtilisin. Therefore, for the subtilisin variant N62D/Y104/G166D the  
25 prosequence would be altered to contain basic residues at positions -4, -2, and -1.

A "signal sequence" or "prosequence" refers to any sequence of amino acids bound to the N-terminal portion of a subtilisin or to the N-terminal portion of a prosubtilisin which may participate in the secretion of the mature or pro forms of the subtilisin. This definition of signal sequence is a functional one, meant to include all those amino acid sequences, encoded by the N-terminal portion of the subtilisin gene or other secretable carbonyl  
30 hydrolases, which participate in the effectuation of the secretion of subtilisin or other carbonyl hydrolases under native conditions. The present invention utilizes such sequences to effect the secretion of the subtilisin mutants as defined herein.

A "prepro" form of a subtilisin mutant consists of the mature form of the subtilisin having a prosequence operably linked to the amino-terminus of the subtilisin and a "pre" or "signal" sequence operably linked to the  
35 amino terminus of the prosequence.

"Expression vector" refers to a DNA construct containing a DNA sequence which is operably linked to a suitable control sequence capable of effecting the expression of the DNA in a suitable host. Such control sequences include a promoter to effect transcription, an optional operator sequence to control such transcription, a sequence encoding suitable mRNA ribosome binding sites, and sequences which control termination of

WO 96/27671

transcription and translation. The vector may be a plasmid, a phage particle, or simply a potential genomic insert. Once transformed into a suitable host, the vector may replicate and function independently of the host genome, or may, in some instances, integrate into the genome itself. In the present specification, "plasmid" and "vector" are sometimes used interchangeably as the plasmid is the most commonly used form of vector at present.

5 However, the invention is intended to include such other forms of expression vectors which serve equivalent functions and which are, or become, known in the art.

The "host cells" used in the present invention generally are prokaryotic or eucaryotic hosts which preferably have been manipulated by the methods disclosed in EPO Publication No. 0130756 or 0251446 or U.S. Patent No. 5,371,008 to render them incapable of secreting enzymatically active endoprotease. A preferred host cell for expressing subtilisin is the *Bacillus* strain BG2036 which is deficient in enzymatically active neutral protease and alkaline protease (subtilisin). The construction of strain BG2036 is described in detail in EPO Publication No. 0130756 and further described by Yang, M.Y., et al. (1984) J. Bacteriol. 160:15-21. Such host cells are distinguishable from those disclosed in PCT Publication No. 03949 wherein enzymatically inactive mutants of intracellular proteases in *E. coli* are disclosed. Other host cells for expressing subtilisin include *Bacillus subtilis* var. II68 (EPO Publication No. 0130756).

15 Host cells are transformed or transfected with vectors constructed using recombinant DNA techniques. Such transformed host cells are capable of either replicating vectors encoding the subtilisin mutants or expressing the desired subtilisin mutant. In the case of vectors which encode the pre or prepro form of the subtilisin mutant, such mutants, when expressed, are typically secreted from the host cell into the host cell medium.

20 "Operably linked" when describing the relationship between two DNA regions simply means that they are functionally related to each other. For example, a presequence is operably linked to a peptide if it functions as a signal sequence, participating in the secretion of the mature form of the protein most probably involving cleavage of the signal sequence. A promoter is operably linked to a coding sequence if it controls the transcription of the sequence; a ribosome binding site is operably linked to a coding sequence if it is positioned 25 so as to permit translation.

25 The genes encoding the naturally-occurring precursor subtilisin may be obtained in accord with the general methods described in U.S. Patent No. 4,760,025 or EPO Publication No. 0130756. As can be seen from the examples disclosed therein, the methods generally comprise synthesizing labeled probes having putative sequences encoding regions of the hydrolase of interest, preparing genomic libraries from organisms expressing 30 the hydrolase, and screening the libraries for the gene of interest by hybridization to the probes. Positively hybridizing clones are then mapped and sequenced.

35 The cloned subtilisin is then used to transform a host cell in order to express the subtilisin. The subtilisin gene is then ligated into a high copy number plasmid. This plasmid replicates in hosts in the sense that it contains the well-known elements necessary for plasmid replication: a promoter operably linked to the gene in question (which may be supplied as the gene's own homologous promotor if it is recognized, i.e., transcribed, by the host), a transcription termination and polyadenylation region (necessary for stability of the mRNA transcribed by the host from the hydrolase gene in certain eucaryotic host cells) which is exogenous or is supplied by the endogenous terminator region of the subtilisin gene and, desirably, a selection gene such as an antibiotic resistance gene that enables continuous cultural maintenance of plasmid-infected host cells by growth in

antibiotic-containing media. High copy number plasmids also contain an origin of replication for the host thereby enabling large numbers of plasmids to be generated in the cytoplasm without chromosomal limitations. However, it is within the scope herein to integrate multiple copies of the subtilisin gene into host genome. This is facilitated by prokaryotic and eukaryotic organisms which are particularly susceptible to homologous recombination.

Once the subtilisin gene has been cloned, a number of modifications are undertaken to enhance the use of the gene beyond synthesis of the naturally-occurring precursor subtilisin. Such modifications include the production of recombinant subtilisin as disclosed in U.S. Patent No. 5,371,008 or EPO Publication No. 0130756 and the production of subtilisin mutants described herein.

10 *Mutant design and preparation.*

A. Subtilisin Variants Capable of Cleaving Substrates Having Dibasic Residues.

For the preparation of subtilisin variants capable of cleaving substrates containing dibasic residues, the following analysis was undertaken.

A number of structures have been solved of subtilisin with a variety of inhibitors and transition state analogs bound (Wright, C. S., Alden, R. A. and Kraut, J. (1969) *Nature*, 221:235-242; McPhalen, C.A. and James, N.G. (1988) *Biochemistry*, 27:6582-6598; Bode, W., Papamokos, E., Musil, D., Seemueller, U. and Fritz, M. (1986) *EMBO J.*, 5:813-818; and Bott, R., Utsch, M., Kossiakoff, A., Graycar, T., Katz, B. and Power, S. (1988) *J. Biol. Chem.*, 263:7895-7906). One of these structures, Figure 1, was used to locate residues that are in close proximity to side chains at the P1 and P2 positions from the substrate. Previous work had shown that replacement residues at positions 156 and 166 in the S1 binding site with various charged residues lead to improved specificity for complementary charged substrates (Wells, J. A., Powers, D. B., Bott, R. R., Graycar, T. P. and Estell, D. A. (1987) *Proc. Natl. Acad. Sci. USA*, 84:1219-1223). Although longer range electrostatic effects of substrate specificity have been noted (Russell, A. J. and Fersht, A. R. (1987) *Nature*, 328:496-500) these were generally much smaller than local ones. Therefore, it seemed reasonable that local differences in charge between subtilisin BPN' and the eukaryotic enzymes may account for the differences in specificity.

A detailed sequence alignment of 35 different subtilisin-like enzymes (Siezen, R. J., de Vos, W. M., Leunissen, A. M., and Dijkstra, B. W. (1991) *Prot. Eng.*, 4:719-737) allowed us to identify differences between subtilisin BPN' and the eukaryotic processing enzymes, KEX2, furin and PC2. Within the S1 binding pocket there are a number of charged residues that appear in the pro-hormone processing enzymes and not in subtilisin BPN' (Table 1A).

TABLE 1A

*SI subsite*

	125-131*	151-157	163-168
Subtilisin BPN'	S L G G P S G (SEQ ID NO: 3)	A A A G N E G (SEQ ID NO: 4)	S T - V G Y P (SEQ ID NO: 5)
Kex2	S W G P A D D (SEQ ID NO: 6)	F A S G N G G (SEQ ID NO: 7)	C N Y D G Y T (SEQ ID NO: 8)
Furin	S W G P E D D (SEQ ID NO: 9)	W A S G N G G (SEQ ID NO: 10)	C N C D G Y T (SEQ ID NO: 11)
PC2	S W G P A D D (SEQ ID NO: 6)	W A S G D G G (SEQ ID NO: 12)	C N C D G Y A (SEQ ID NO: 13)

\* numbering according to subtilisin BPN' sequence

For example, the eukaryotic enzymes have two conserved Asp residues at 130 and 131 as well as an Asp at 165  
10 that is preceded by insertion of a Tyr or Cys. However, in the region from 151-157, subtilisin BPN' contains a Glu and the eukaryotes a conserved Gly.

In the S2 binding site there were two notable differences in sequence (Table 1B).

TABLE 1B

*S2 subsite*

	30-35	60-64
Subtilisin BPN'	V I D S G I (SEQ ID NO: 14)	D N N S H (SEQ ID NO: 15)
KEX2	I V D D G L (SEQ ID NO: 16)	S D D Y H (SEQ ID NO: 17)
Furin	I L D D G I (SEQ ID NO: 18)	N D N R H (SEQ ID NO: 19)
PC2	I M D D G I (SEQ ID NO: 20)	W F N S H (SEQ ID NO: 21)

Subtilisin BPN' contains a Ser at position 33 whereas the pro-hormone processing enzymes contain Asp. There  
20 is not as clear a consensus in the region of 60-64, but one notable difference is at position 62. This side chain which points directly at the P2 side chain (Figure 1) is Asn in subtilisin BPN', furin and PC2 but Asp in KEX2. Thus, not all substitutions were clearly predictive of the specificity differences.

A variety of mutants were produced to probe and engineer the specificity of subtilisin BPN' using oligonucleotides described in Table 2.

TABLE 2  
*Oligonucleotides used for site-directed  
 mutagenesis on subtilisin.*

Mutant	Oligonucleotide	Specificity Pocket	Activity Expressed
S33D	5'- GCGGTTATCGACG°A°CGGTATCGATTCT -3' (SEQ ID NO: 22)	S2	+
S33K	5'- GCGGTTATCGACAA°A°G°GTATCGATTCT -3' (SEQ ID NO: 23)	S2	+
S33E	5'- GCGGTTATCGACG°A°A°GGTATCGATTCT -3' (SEQ ID NO: 24)	S2	+
N62D	5'- CCAAGACAACG°ACTCTCACGGAA -3' (SEQ ID NO: 25)	S2	+
N62S	5'- CCAAGACAACAG°CTCTCACGGAA -3' (SEQ ID NO: 26)	S2	+
N62K	5'- CCAAGACAACAAA°TCTCACGGAA -3' (SEQ ID NO: 27)	S2	+
G166D	5'-CACTCCGGCAGCTCG°T°C°G°ACAGTGGAA°C°T ACCCGGC.AAATA-3' (SEQ ID NO: 28) (Inserts Sal I site)	S1	+
G166E	5'-CACTCCGGCAGCTCG°T°C°G°ACAGTGGAA°GT ACCCGGCAAATA-3' (SEQ ID NO: 29) (Inserts Sal I site)	S1	+
G128P/P129A	5'- TTAACATGAGCCTCGGCC°C°AG°CTA°G°C°GGT TCTGCTGCTTTA -3' (SEQ ID NO: 30) (Inserts Nhe I site)	S1	-
G128P/P129A/ S130D/G131D	5'- TTAACATGAGCCTCGGCC°C°C°G°CGG°A°TGA° TTCTGCTGCTTTAA -3' (SEQ ID NO: 31) (Inserts Sac II site)	S1	-
T164N/V165D	5'-CGGCAGCTCAAGCA°A°C°G°A°T°GGCTAT°CCT GGCAAATACCCCTCTGTCA -3' (SEQ ID NO: 32) (Inserts BsaBI site)	S1	-

T164Y/V165D	5'-CGGCAGCTCAAGCA*A*C*G*A*T*GGCTAT*CCT GGCAAATACCCCTCTGTCA -3' (SEQ ID NO: 33) (Inserts BsaBI site)	S1	-
T164N-Y(insert)-V165D	5'-ACTTCCGGCAGCTCT*T*C*G*AA*C*T*A*C*G*A* C*GGGTACCCGGCAAATA-3' (SEQ ID NO: 34) (Inserts BstBI site)	S1	-
N62D/G166D	See individual mutations	S1/S2	-
N62D/G166E	See individual mutations	S1/S2	+

\* Asterisks indicate base changes from the pSS5 (wild-type) template.

After producing the mutant plasmids they were transformed into a protease deficient strain of *B. subtilis* (BG2036) that lacks an endogenous gene for secretion of subtilisin. These were then tested for protease activity on skim milk plates.

The first set of mutants tested were ones where segments of the S1 binding site were replaced with sequences from KEX2. None of these segment replacements produced detectable activity on skim milk plates even though variants of subtilisin whose catalytic efficiencies are reduced by as much as 1000-fold do produce detectable halos (Wells, J.A., Cunningham, B.C., Graycar, T.P. and Estell, D.A. (1986) *Philos. Trans. R. Soc. Lond. A.* 317:415-423). We went on to produce single residue substitutions that should have less impact on the stability. These mutants at positions 166 in the S1 site, and 33 and 62 in the S2 site, were chosen based on the modeling and sequence considerations described above. Fortunately all single mutants as well as combination mutants produced activity on skim milk plates and could be purified to homogeneity.

#### 20 Kinetic analysis of variant subtilisins.

To probe the effects of the G166E and G166D on specificity at the P1 position we used substrates having the form suc-AAPX-pna (SEQ ID NO: 69) where X was either Lys (SEQ ID NO. 58), Arg (SEQ ID NO. 59), Phe (SEQ ID NO. 56), Met (SEQ ID NO. 60) or Gln (SEQ ID NO. 61). The  $k_{cat}/K_m$  values were determined from initial rate measurements and results reported in Figure 2. Whereas the wild-type enzyme preferred Phe>Met>Lys>Arg>Gln, the G166E preferred Lys>Phe>Arg>Met>Gln, and G166D preferred Lys>Phe>Arg>Met>Gln. Thus, both the acidic substitutions at position 166 caused a shift in preference for basic residues at the P1 site, as previously reported (Wells, J. A., Powers, D. B., Bott, R. R., Graycar, T. P. and Estell, D. A. (1987a), *Proc. Natl. Acad. Sci. USA* 84:1219-1223).

The effects of single and double substitutions in the S2 binding site were analyzed with substrates having the form, suc-Ala-Ala-Xaa-Phe-pna (SEQ ID NO. 70) and are shown in Figure 3. At the P2 position the wild-type enzyme preferred Ala>Pro>Lys>Arg>Asp. In contrast, the S33D preferred Ala>Lys>Arg>Pro>Asp and the N62D preferred Lys>Ala>Arg>Pro>Asp. Although the effects were more dramatic for the N62D mutant, the S33D variant also showed significant improvement toward basic P2 residues and corresponding reduction

in hydrolysis of the Ala and Asp P2 substrates. We then analyzed the double mutant, but found it exhibited the catalytic efficiency of the worse of the two single mutants for each of the substrates tested.

Despite the less than additive effects seen for the two charged substitutions in the S2 site, we decided to combine the best S2 site variant (N62D) with either of the acidic substitutions in the S1 site. The two double mutants, N62D/G166E and N62D/G166D, were analyzed with substrates having the form, suc-AAXX-pna (SEQ ID NO. 71) where XX was either KK (SEQ ID NO. 66), KR (SEQ ID NO. 67), KF (SEQ ID NO. 62), PK (SEQ ID NO. 58), PF (SEQ ID NO. 56) or AF (SEQ ID NO. 63) (Figure 4). The wild-type preference was AF>PF>KF>KK>PK>KR, whereas the double mutants had the preference KK>KR>KF>PK>AF>PF. Thus for the double mutants there was a dramatic improvement toward cleavage of dibasic substrates and away from cleaving the hydrophobic substrates.

The greater than additive effect (or synergy) of these mutants can be seen from ratios of the catalytic efficiencies for the single and multiple mutants. For example, the G166E variant cannot distinguish Lys from Phe at the P1 position. Yet the N62D/G166E variant cleaves the Lys-Lys substrate about 8 times faster than the Lys-Phe substrate. Similarly the G166D cleaves the Lys P1 substrate about 3 times faster than the Phe P1 substrate, but the N62D/G166D double mutant cleaves a Lys-Lys substrate 18 times faster than a Lys-Phe substrate. Thus, as opposed to the reduction in specificity seen for the double mutant in the S2 site, the S1-S2 double mutants enhance specificity for basic residues. It is possible that these two sites bind the dibasic substrates in a cooperative manner analogous to a chelate effect.

Therefore, according to the present invention, subtilisin mutants having a preference for dibasic residues are preferred. According to this aspect of the present invention substitution of amino acids corresponding to amino acids N62 and G166 of subtilisin BPN' produced from *Bacillus amyloliquefaciens* are prepared. In particular, amino acids 62 and 166, or their equivalents, in the precursor subtilisin are substituted with amino acid residues Asp or Glu. Preferred subtilisin variants according to this aspect of the invention include N62D/G166D, N62E/G166E, N62E/G166D, and N62D/G166E variants of subtilisin BPN' and their equivalents.

25      B. Subtilisin Variants Capable of Cleaving Substrates Having Tribasic Residues

For the preparation of subtilisin variants specific for substrates containing a third basic residue at substrate position P4 we used the crystal structure of subtilisin BPN' complexed with Ala-Ala-Pro-Phe-Boronate (SEQ ID NO: 56) (Figure 7) in combination with sequence alignments of subtilisin BPN', KEX2, Furin, PC2, and P (Table 3) in designing basic specificity into the S1 and S2 and S4 subsites. The two subtilisin BPN' residues that most prominently display their side chains into the S4 pocket are Y104 and I107 (Figure 7).

Sequence alignments of subtilisin BPN' and the mammalian prohormone-processing proteases (Siezen, R. J., de Vos, W. M., Leunissen, A. M., and Dijkstra, B. W. (1991) *Prot. Eng.* 4:719-737) (Table 3) reveal that position 104 is conserved as Asp, and 107 as Glu in the prohormone converting (Arg-P4 specific) enzymes. Therefore these two mutations were introduced either individually or in combination into the dibasic-specific

35      N62D/G166D subtilisin BPN' background (Table 4).

Table 3 Sequence alignments for the S4 site of subtilisins

	S4 Site	
	100-110	
5	Subtilisin	GSGQYSWIING (SEQ ID NO: 77)
	KEX2	GDITTEDEAAS (SEQ ID NO: 78)
	Furin	GEVTDAAEVEARS (SEQ ID NO: 79)
	PC2	PPMTDIIEASS (SEQ ID NO: 80)
	P	GIVTDAIEASS (SEQ ID NO: 81)

10 Table 4 describes oligonucleotides used for site-directed mutagenesis, protein regions affected by the mutations, and relative expression of protein for N62D/G166D subtilisin BPN' variants. Bold type indicates base changes from the pSS5 (N62D/G166D) template. For "Protein Expressed," "+" indicates a high level of expression of mature enzyme in crude culture medium, and "-" indicates no enzyme detectable.

TABLE 4

	Oligonucleotide	Protein Region	Protein Expressed
15	Y104D	5'- GGTCGGGCCAA <b>GATAGCTGGATCATT</b> -3' (SEQ ID NO: 82)	S4 pocket -
	I107E	5'- CCAATACAGCTGGAAATTAAACGGAATCG -3' (SEQ ID NO: 83)	S4 pocket -
	Y104D/I107E	5'- GGTCGGCCAA <b>GAGATAGCTGGAAATTAAACG</b> GAATCGA -3' (SEQ ID NO: 84)	S4 pocket -
20	A(-4)R/ A(-2)K/ Y(-1)R	5'- AAGAAGATCACGTAAAGACATAAGCGCGCG AGTCGTGC -3' (SEQ ID NO: 85)	Processing site -
	Y104D/ A(-4)R/ A(-2)K/ Y(-1)R	See individual mutations	S4 pocket + Processing site +
25	I107E/ A(-4)R/ A(-2)K/ Y(-1)R	See individual mutations	S4 pocket + Processing site +
30	Y104D/I107E/ A(-4)R/ A(-2)K/ Y(-1)R	See individual mutations	S4 pocket + Processing site +

Initial attempts to express the triple mutants in *Bacillus* were unsuccessful, as indicated by SDS-PAGE of crude supernatants. We reasoned that the source of the expression problem could lie in the fact that correct folding and maturation of subtilisin requires autolytic cleavage of its propeptide (Power, S.D., Adams, R. M., and Wells, J. A. (1986) *Proc. Natl. Acad. Sci. USA* 83, 3096-3100). The processing site in 5 the wild-type enzyme has a sequence that is optimized for the natural substrate preference, AHAY<sup>1</sup>A (1 denotes the site of cleavage). Although the N62D/G166D subtilisin can still autolyze itself with the wild-type processing site, the additional S4 pocket mutations could reduce the cleavage to the point where expression was lowered to a minute level.

To test whether the mutants were expressed poorly due to an inability to autolytically process itself, 10 mutations in the processing site were simultaneously incorporated to accommodate the changes in substrate specificity. Thus the sequence from positions -4 to -1 was changed from AHAY to RHKR in combination with the S4 site mutations. For N62D/Y104D/G166D, high levels of expression could then be achieved providing an indication that the additional Y104D mutation induced an especially strong preference for P4 15 Arg over Ala. Variants containing the I107E mutation, however, could not be expressed even with the change in the processing site.

#### Kinetic analysis of variant subtilisins

The mature N62D/Y104D/G166D variant was purified and analyzed for its ability to hydrolyze several tetrapeptide-pNA substrates. Table 5 displays the results along with data for the N62D/G166D mutant and wild-type subtilisin.

Table 5

Kinetic analysis of WT, N62D/G166D, and N62D/Y104D/G166D mutants versus succinyl-tetrapeptide-pNA substrates. Kinetic constants were determined from plots of initial rates versus substrate concentration. Units are as follows:  $k_{cat}$ , s<sup>-1</sup>; Km,  $\mu$ M; and  $k_{cat}/Km$ , M<sup>-1</sup>s<sup>-1</sup>. Standard errors were less than 15%.

		Tetrapeptide Sequence, P4-P3-P2-P1						RAKR				RARR				
		AAPF			AAKR			RAKR			RARR			RARR		
Mutant	$k_{cat}$	Km	$k_{cat}/Km$	$k_{cat}$	Km	$k_{cat}/Km$	$k_{cat}$	Km	$k_{cat}/Km$	$k_{cat}$	Km	$k_{cat}/Km$	$k_{cat}$	Km	$k_{cat}/Km$	
WT	29	110	$2.6 \times 10^3$	2.8	1700	$1.7 \times 10^3$	N.D.	N.D.	N.D.	N.D.	N.D.	N.D.	N.D.	N.D.	N.D.	
N62D/ G166D	3.4	180	$1.9 \times 10^3$	15	41	$3.7 \times 10^3$	N.D. <sup>b</sup>	N.D. <sup>b</sup>	N.D. <sup>b</sup>	N.D. <sup>b</sup>	N.D.	N.D.	N.D.	N.D.	N.D.	
N62D/ Y104D/ G166D	-	-	$8.0 \times 10^6$	0.27	34	$8.1 \times 10^3$	11	23	$4.9 \times 10^5$	9.8	29	$3.4 \times 10^5$	-	-	-	

- <sup>a</sup> Artificially low  $k_{cat}/Km$  presumably due to competing cleavage between Lys and Arg
- <sup>b</sup> Biphasic reaction progress curves presumably due to multiple cleavages in the substrate
- <sup>c</sup> Unable to saturate the enzyme, apparent  $k_{cat}/Km$  calculated from rates at low substrate concentrations assuming  $v = (k_{cat}/Km) [E] (S)$

The tribasic substrates succinyl-RAKR-pNA (SEQ ID NO: 86) and succinyl-KAKR-pNA (SEQ ID NO: 87) were hydrolyzed with high catalytic efficiency ( $k_{cat}/K_m$ ) by the triple mutant, at a level similar to wild-type subtilisin versus one of its best substrates, succinyl-AAPF-pNA (SEQ ID NO: 56). In contrast, the dibasic substrate succinyl-AAKR-pNA (SEQ ID NO: 67) was hydrolyzed 60-fold less efficiently, mostly

5 due to diminution of  $k_{cat}$ . This indicates a dramatic specificity change from the wild-type preference at P4, at which hydrophobic residues are strongly favored over charged side chains (Gren, H. and Breddam, K. (1992) *Biochemistry* 31, 8967-8971). In fact N62D/G166D subtilisin appears to cleave at an alternate site in the succinyl-RAKR-pNA (SEQ ID NO: 86) substrate, indicating that Arg was not accepted in its wild-type S4 site.

10 The large magnitude of the combined specificity changes in the N62D/Y104D/G166D variant is evidenced by its strong discrimination against substrates that are preferred by the wild-type enzyme. For example, succinyl-AAPF-pNA (SEQ ID NO: 56) is hydrolyzed  $6 \times 10^4$ -fold less efficiently than succinyl-RAKR-pNA (SEQ ID NO: 86). Clearly, the S4 site mutation greatly improves upon the discriminatory power of the parent dibasic-specific N62D/G166D subtilisin, where the ratio of catalytic efficiency for 15 succinyl-AAKR-pNA versus succinyl-AAPF-pNA is  $1.9 \times 10^4$ . The improvement in discrimination (310-fold) is also higher than would be predicted from the data for hydrolysis of succinyl-RAKR-pNA (SEQ ID NO: 86) versus succinyl-AAKR-pNA (SEQ ID NO: 67) by the triple mutant (a 60-fold effect).

Therefore in order to produce subtilisin variants capable of cleaving substrates containing basic residues at positions P4, P2, and P1, additional site specific substitutions are made in the dibasic specific 20 subtilisin variants. According to this aspect of the invention, substitution of the amino acid corresponding to Y104 of subtilisin BPN' produced by *Bacillus Amyloliquefaciens*, i.e., amino acid 104 of subtilisin BPN' or its equivalent, produces a variant having substantially altered substrate specificity. In a preferred embodiment of the present invention amino acids corresponding to N62, Y104, and G166 of subtilisin BPN' are substituted with acidic amino acids, preferably Asp and Glu and most preferably Asp. Subtilisin BPN' 25 variants N62D/Y104D/G166D, N62D/Y104E/G166D, N62E/Y104D/G166E, N62E/Y104E/G166E, N62E/Y104D/G166D, N62E/Y104E/G166D, N62D/Y104E/G166E, and N62D/Y104D/G166E, and their equivalents are preferred. Most preferred among this group of subtilisin variants are the N62D/Y104D/G166D subtilisin BPN' variants and their equivalents.

#### Mutagenesis and Synthetic Techniques

30 Various techniques are available which may be employed to produce mutant DNA, which can encode the subtilisin variants of the present invention. For instance, it is possible to derive mutant DNA based on naturally occurring DNA sequences that encode for changes in an amino acid sequence of the resultant protein relative to a precursor subtilisin. These mutant DNA can be used to obtain the variants of the present invention.

35 According to the invention, specific residues of *B. amyloliquefaciens* subtilisin are identified for substitution. These amino acid residue position numbers refer to those assigned to the *B. amyloliquefaciens* subtilisin sequence (see the mature sequence in Fig. 1. of U.S. Patent No. 4,760,025). The invention, however, is not limited to the mutation of this particular subtilisin but extends to precursor subtilisins

containing amino acid residues which are equivalent, as defined herein, to the particular identified residues in *B. amyloliquefaciens* subtilisin. Equivalent amino acids can be found in, for instance, subtilisin Carlsberg from *Bacillus licheniformis*, Bode et al., (1986) EMBO J., 5:813-818; thermitase from *Thermoactinomyces vulgaris*, Gros et al., (1989) J. Mol. Biol. 210:347-367; and proteinase K from *Tritirachium album*, Betzel, et al., (1988) Acta Crystlogr., B, 44:163-172) as described by Siezen et al., (1991) Prof. Eng., 4: 719-737).

By way of illustration, with expression vectors encoding the precursor subtilisin in hand (see for example U.S. Patent No 4,760,025) site specific mutagenesis (Kunkel et al., (1991) Methods Enzymol. 204:125-139; Carter, P., et al., (1986) Nucl. Acids. Res. 13:4331; Zoller, M. J. et al., (1982) Nucl. Acids Res. 10:6487), cassette mutagenesis (Wells, J. A., et al., (1985) Gene 34:315), restriction selection 10 mutagenesis (Wells, J. A., et al., (1986) Philos. Trans. R. Soc. London Ser A 317, 415) or other known techniques may be performed on the DNA. The mutant DNA can then be used in place of the parent DNA by insertion into the appropriate expression vectors. Growth of host bacteria containing the expression vectors with the mutant DNA allows the production of variants which can be isolated as described herein.

Oligonucleotide-mediated mutagenesis is a preferred method for preparing the variants of the present 15 invention. This technique is well known in the art as described by Adelman et al., (1983) DNA. 2:183. Briefly, the native or unaltered DNA of a precursor subtilisin, for instance subtilisin BPN', is altered by hybridizing an oligonucleotide encoding the desired mutation to a DNA template, where the template is the single-stranded form of a plasmid or bacteriophage containing the unaltered or native DNA sequence of the precursor.

20 After hybridization, a DNA polymerase is used to synthesize an entire second complementary strand of the template that will thus incorporate the oligonucleotide primer, and will code for the selected alteration in the DNA.

Generally, oligonucleotides of at least 25 nucleotides in length are used. An optimal oligonucleotide 25 will have 12 to 15 nucleotides that are completely complementary to the template on either side of the nucleotide(s) coding for the mutation. This ensures that the oligonucleotide will hybridize properly to the single-stranded DNA template molecule. The oligonucleotides are readily synthesized using techniques known in the art such as those described by Crea et al. (1987) Proc. Natl. Acad. Sci. USA, 75:5765. Exemplary oligonucleotide sequences for introducing amino acid changes into precursor subtilisin BPN' are provided in Tables 2 and 4.

30 Single-stranded DNA template may also be generated by denaturing double-stranded plasmid (or other) DNA using standard techniques.

For alteration of the native DNA sequence (to generate amino acid sequence variants, for example), the oligonucleotide is hybridized to the single-stranded template under suitable hybridization conditions. A DNA polymerizing enzyme, usually the Klenow fragment of DNA polymerase I, is then added to 35 synthesize the complementary strand of the template using the oligonucleotide as a primer for synthesis. A heteroduplex molecule is thus formed such that one strand of DNA encodes the variant form of the subtilisin, and the other strand (the original template) encodes the native, unaltered sequence of the precursor subtilisin. This heteroduplex molecule is then transformed into a suitable host cell. After the cells are grown, they are plated onto agarose plates and screened using the oligonucleotide primer radiolabeled with 32-phosphate to

identify the bacterial colonies that contain the mutated DNA. The mutated region is then removed and placed in an appropriate vector for protein production, generally an expression vector of the type typically employed for transformation of an appropriate host.

The method described immediately above may be modified such that a homoduplex molecule is created wherein both strands of the plasmid contain the mutation(s). The modifications are as follows: The single-stranded oligonucleotide is annealed to the single-stranded template as described above. A mixture of three deoxyribonucleotides, deoxyriboadenosine (dATP), deoxyriboguanosine (dGTP), and deoxyribothymidine (dTTP), is combined with a modified thio-deoxyribocytosine called dCTP-( $\alpha$ S) (which can be obtained from Amersham Corporation). This mixture is added to the template-oligonucleotide complex. Upon addition of DNA polymerase to this mixture, a strand of DNA identical to the template except for the mutated bases is generated. In addition, this new strand of DNA will contain dCTP-( $\alpha$ S) instead of dCTP, which serves to protect it from restriction endonuclease digestion.

After the template strand of the double-stranded heteroduplex is nicked with an appropriate restriction enzyme, the template strand can be digested with ExoIII nuclease or another appropriate nuclease past the region that contains the site(s) to be mutagenized. The reaction is then stopped to leave a molecule that is only partially single-stranded. A complete double-stranded DNA homoduplex is then formed using DNA polymerase in the presence of all four deoxyribonucleotide triphosphates, ATP, and DNA ligase. This homoduplex molecule can then be transformed into a suitable host cell as described above.

DNA encoding variants with more than one amino acid to be substituted may be generated in one of several ways. If the amino acids are located close together in the polypeptide chain, they may be mutated simultaneously using one oligonucleotide that codes for all of the desired amino acid substitutions. If, however, the amino acids are located some distance from each other (separated by more than about ten amino acids), it is more difficult to generate a single oligonucleotide that encodes all of the desired changes. Instead, one of two alternative methods may be employed.

In the first method, a separate oligonucleotide is generated for each amino acid to be substituted. The oligonucleotides are then annealed to the single-stranded template DNA simultaneously, and the second strand of DNA that is synthesized from the template will encode all of the desired amino acid substitutions.

The alternative method involves two or more rounds of mutagenesis to produce the desired mutant. The first round is as described for the single mutants: wild-type DNA is used for the template, an oligonucleotide encoding the first desired amino acid substitution(s) is annealed to this template, and the heteroduplex DNA molecule is then generated. The second round of mutagenesis utilizes the mutated DNA produced in the first round of mutagenesis as the template. Thus, this template already contains one or more mutations. The oligonucleotide encoding the additional desired amino acid substitution(s) is then annealed to this template, and the resulting strand of DNA now encodes mutations from both the first and second rounds of mutagenesis. This resultant DNA can be used as a template in a third round of mutagenesis, and so on.

*Cleavage of a Fusion Protein With Subtilisin Variants*

A fusion protein is any polypeptide that contains within it an affinity domain (AD) that usually aids in protein purification, a protease cleavage sequence or substrate linker (SL), which is cleaved by a protease and a protein product of interest (PP). Such fusion proteins are generally expressed by recombinant DNA technology. The genes for fusion proteins are designed so that the SL is between the AD and PP. These usually take the form AD-SL-PP such that the domain closest to the N-terminus is AD and PP is closest to the C-terminus.

Examples of AD would include, glutathione-S-transferase which binds to glutathione, protein A (or derivatives or fragments thereof) which binds IgG molecules, poly-histidine sequences, particularly (His)<sub>6</sub> (SEQ ID NO: 51) that bind metal affinity columns, maltose binding protein that binds maltose, human growth hormone that binds the human growth hormone receptor or any of a variety of other proteins or protein domains that can bind to an immobilized affinity support with an association constant ( $K_a$ ) of  $>10^5$  M<sup>-1</sup>.

The SL can be any sequence which is cleaved by the subtilisin variants of the present invention. In preparations where the variant N62D/Y104D/G166D or its equivalent are used the SL can be any sequence, preferably at least 4 amino acids, in which the P4, P2, and P1 residues are basic residues. Therefore a SL linker is employed of the general formula P4-P3-P2-P1 wherein P4, P2, and P1 are basic amino acid residues. Preferred SLs according to this aspect of the invention include Lys-Ala-Lys-Arg (SEQ ID NO: 87) and Arg-Ala-Lys-Arg (SEQ ID NO: 86).

Likewise, where the N62D/G166D subtilisin variant is contemplated the SL preferably contains di-basic residues. For the variants capable of cleaving substrates containing dibasic residues the SL should be at least four residues and preferably contain a large hydrophobic residue at P4 (such as Leu or Met) and dibasic residues at P2 and P1 (such as Arg and Lys). A particularly good substrate is Leu-Met-Arg-Lys- (SEQ ID NO: 52), but a variety of other sequences may work including Ala-Ser-Arg-Arg (SEQ ID NO: 50) and even Leu-Thr-Ala-Arg (SEQ ID NO 53).

It is often useful that the SL contain a flexible segment on its N-terminus to better separate it from the AD and PP. Such sequences include Gly-Pro-Gly-Gly (SEQ ID NO: 54) but can be as simple as Gly-Gly or Pro-Gly. Thus, an example of a particularly good SL would have the sequence Gly-Pro-Gly-Gly-Leu-Met-Arg-Lys (SEQ ID NO: 88) in the case of subtilisin variants capable of cleaving substrates containing dibasic amino acids, or Gly-Pro-Gly-Gly-Ala-Lys-Arg (SEQ ID NO: 89). This sequence would be inserted between the AD and PP domains.

The PP can be virtually any protein or peptide of interest but preferably should not have a Pro, Ile, Thr, Val, Asp or Glu as its first residue (P1'), or Pro or Gly at the second residue (P2') or Pro at the third residue (P3'). Such residues are poor substrates for the enzyme and may impair the ability of the subtilisins variant to cleave the SL sequence.

The conditions for cleaving the fusion protein are best done in aqueous solution, although it should be possible to immobilize the enzyme and cleave the soluble fusion protein. It may also be possible to cleave the fusion protein as it remains immobilized on a solid support (e.g. bound to the solid support through AD) with the soluble subtilisin variant. It is preferable to add the enzyme to the fusion protein so that the enzyme

is less than one part in 100 (1:100) by weight. A good buffer is 10-50mM Tris (pH 8.2) in 10mM NaCl. A preferable temperature is about 25°C although the enzyme is active up to 65°C. The extent of cleavage can be assayed by applying samples to SDS-PAGE. Generally suitable conditions for using the subtilisin variants of this invention do not depart substantially from those known in the art for the use of other subtilisins.

5

#### EXAMPLES

In the examples below and elsewhere, the following abbreviations are employed: subtilisin BPN', subtilisin from *Bacillus amyloliquefaciens*; Boc-RVRR-MCA (SEQ ID NO. 73), N-t-butoxy carbonyl-arginine-valine-arginine-arginine-7-amido-4-methyl coumarin; suc-Ala-Ala-Pro-Phe-pna (SEQ ID NO. 56), N-succinyl-alanine-alanine-proline-phenylalanyl-p-nitroanalide (SEQ ID NO. 56); hGH, human growth hormone; hGHbp, extracellular domain of the hGH receptor; PBS, phosphate buffered saline; AP, alkaline phosphatase;

10

10

#### Example 1

##### *Construction and Purification of Subtilisin Mutants.*

15 Site-directed mutations were introduced into the subtilisin BPN' gene cloned into the phagemid pSS5 (Wells, J. A., Ferrari, E., Henner, D. J., Estell, D.A. and Chen, E. Y. (1983) *Nucl. Acids Res.* 11:7911-7929). Single-stranded uracil-containing pSS5 template was prepared and mutagenesis performed using the method of Kunkel (Kunkel, T. A., Bebenek, K and McClary, J. (1991) *Methods Enzymol.* 204:125-139). For example, the synthetic oligonucleotide N62D,

20

(5'-CCAAGACAACG<sup>\*</sup>ACTCTCACGGAA -3') (SEQ ID NO. 25)

in which the asterisk denotes a mismatch to the wild-type sequence, was used to construct the N62D mutant. The oligonucleotide was first phosphorylated at the 5' end using T4 polynucleotide kinase according to a described procedure (Sambrook, J., Fritsch, E. F., and Maniatis, T. (1989) in "Molecular Cloning: A Laboratory Manual," Second Edition, Cold Spring Harbor, N.Y.). The phosphorylated oligonucleotide was annealed to single-stranded uracil-containing pSS5 template, the complementary DNA strand was filled in with deoxynucleotides using T7 polynucleotide kinase, and the resulting nicks ligated using T4 DNA ligase according to a previously described procedure (Sambrook, J., Fritsch, E. F., and Maniatis, T. (1989) in "Molecular Cloning: A Laboratory Manual," Second Edition, Cold Spring Harbor, N.Y.). Heteroduplex 25

30

DNA was transformed into the *E. coli* host JM101(Yanish-Perron, C., Viera, J., and Messing, J. (1985) *Gene* 33: 103-199), and putative mutants were confirmed by preparation and dideoxy nucleotide sequencing of single stranded DNA (Sanger, F., Nicklen, S. and Coulson, A. R. (1977) *Proc. Natl. Acad. Sci. USA* 74:5463-5467) according to the SEQUENASE<sup>®</sup> protocol (USB Biochemicals). Mutant single-stranded DNA was then retransformed into JM101 cells and double stranded DNA prepared according to a previously

WO 96/27671

described procedure (Sambrook, J.; Fritsch, E. F., and Maniatis, T. (1989) in "Molecular Cloning: A Laboratory Manual," Second Edition, Cold Spring Harbor, N.Y.). For other mutations also requiring the use of one primer, the oligonucleotides used are listed in Table 2. For several of these oligonucleotides, additional silent mutations emplacing new restriction sites were simultaneously introduced to provide an alternative verification of mutagenesis.

To construct the double mutants N62D/G166D, and N62D/G166E, pSS5 DNA containing the N62D mutation was produced in single-stranded uracil-containing form using the Kunkel procedure (Kunkel, T. A., Bebenek, K and McClary, J. (1991) *Methods Enzymol.* 204, 125-139). This mutant DNA was used as template for the further introduction of the G166D or G166E mutations, using the appropriate oligonucleotide primers (see sequences in Table 2), following the procedures described above.

To construct the triple mutants, such as N62D/Y104D/G166D, pSS5 DNA containing the N62D/G166D mutation or other appropriate double mutation, was produced in single-stranded uracil-containing form using the Kunkel procedure (Kunkel, T. A., Bebenek, K and McClary, J. (1991) *Methods Enzymol.* 204, 125-139). This mutant DNA was used as template for the further introduction of the Y104D mutations, using the appropriate oligonucleotide primers (see sequences in Table 4), following the procedures described above.

For expression of the subtilisin BPN' mutants, double stranded mutant DNA was transformed into a protease-deficient strain (BG2036) of *Bacillus Subtilis* (Yang, M. Y., Ferrari, E. and Henner, D. J. (1984) *Journal of Bacteriology* 160:15-21) according to a previous method (Anagnostopoulos, C. and Spizizen, J. (1961) *Journal of Bacteriology* 81:741-746) in which transformation mixtures were plated out on LB plus skim milk plates containing 12.5 µg/mL chloramphenicol. The clear halos indicative of skim milk digestion surrounding transformed colonies were noted to roughly estimate secreted protease activity.

The transformed BG2036 strains were cultured by inoculating 5 mL of 2xYT media (Miller, J. H., (1972) in "Experiments in Molecular Genetics," Cold Spring Harbor, N.Y.) containing 12.5 µg/mL chloramphenicol and 2 mM CaCl<sub>2</sub>, at 37 °C for 18-20 h, followed by 1:100 dilution in the same medium and growth in shake flasks at 37 °C for 18-22 h with vigorous aeration. The cells were harvested by centrifugation (6000g, 15 min, 4°C), and to the supernatant 20mM (final) CaCl<sub>2</sub> and one volume of ethanol (-20°C) were added. After 30 min at 4°C, the solution was centrifuged (12,000g, 15 min, 4°C), and one volume of ethanol (-20°C) added to the supernatant. After 2 h at -20°C, the solution was centrifuged (12,000g, 15 min, 4°C) and the pellet resuspended in and dialyzed against MC (25 mM 2-(N-Morpholino)ethanesulfonic acid (MES), 5 mM CaCl<sub>2</sub> at pH 5.5) overnight at 4°C. The dialysate was passed through a 0.22 µm syringe filter and loaded onto a mono-S cation exchange column run by an FPLC system (Pharmacia Biotechnology). The column was washed with 20 volumes of MC and mutant subtilisin eluted over a linear gradient of zero to 0.15 M NaCl in MC, all at a flow rate of 1 mL/min. Peak fractions were recovered and the subtilisin mutant quantitated by measuring the absorbance at 280 nm ( $E_{280}$  0.1% = 1.17) (Matsubara, H.; Kasper, C B.; Brown, D. M.; and Smith, E. L. (1965) *J. Biol. Chem.*, 240:1125-1130.)

## Example 2

*Kinetic Characterizations*

Subtilisins were assayed by measuring the initial rates of hydrolysis of *p*-nitroanilide tetrapeptide substrates in 0.4 mL 20 mM Tris-Cl pH 8.2, 4% (v/v) dimethyl sulfoxide at (25 ± 0.2)°C as described previously (Estell, D. A., Graycar, T. P., Miller, J. V., Powers, D. B., Burnier, J. P., Ng, P. G. and Wells, J. A. (1986) *Science* 233:659-663). Enzyme concentrations ( $E_0$ ) were determined spectrophotometrically using  $E_{300\text{nm}}=0.1\% \approx 1.17$  (Matsubara, H.; Kasper, C.B.; Brown, D.M.; and Smith, E.L. (1965) *J. Biol. Chem.*, 240:1125-1130.), and were typically 5-50 nM in reactions. Initial rates were determined for nine to twelve different substrate concentrations over the range of 0.001-2.0 mM. Plots of initial rates (v) versus substrate concentration (S) were fitted to the Michaelis-Menton equation,

$$v = \frac{k_{cat}(E_0)(S)}{K_m + S}$$

to determine the kinetic constants  $k_{cat}$  and  $K_m$  (Fersht, A. in "Enzyme Structure and Mechanism", Second edition, Freeman and Co., N.Y.) using the program Kaleidagraph (Synergy Software, Reading, PA).

## Example 3

*Substrate Phage*

Substrate phage selections were performed as described by Matthews and Wells (Matthews, D. J. and Wells, J. A. (1993) *Science* 260:1113-1117), with minor modifications. Phage sorting was carried out using a library in which the linker sequence between the gene III coat protein and a tight-binding variant of hGH was GPGGX<sub>3</sub>GGPG (SEQ ID NO. 52). The library contained  $2 \times 10^6$  independent transformants. Phage particles were prepared by infecting 1 mL of log phase 27C7 (F/*lac<sup>R</sup>/Omp<sup>T</sup>/degP<sup>r</sup>) *Escherichia coli* with approximately  $10^8$  library phage for 1 h at 37°C, followed by 18-24 h of growth in 25 mL 2YT medium containing  $10^{10}$  M13K07 helper phage and 50 µg/mL carbenicillin at 37°C. Wells of a 96-well Nunc Maxisorb microtiter plate were coated with 2 µg/mL of hGHbp in 50 mM NaHCO<sub>3</sub> at pH 9.6 overnight at 4°C and blocked with PBS (10 mM sodium phosphate at pH 7.4 and 150 mM NaCl) containing 2.5% (w/v) skim milk for 1 h at room temperature. Between  $10^{11}$  and  $10^{12}$  phage in 0.1 mL 10 mM tris-Cl (pH 7.6), 1 mM EDTA, and 100 mM NaCl were incubated in the wells at room temperature for 2 h with gentle agitation. The plate was washed first with 20 rinses of PBS plus 0.05% Tween 20 and then twice with 20 mM tris-Cl at pH 8.2. The N62D/G166D subtilisin was added in 0.1 mL of 20 mM tris-Cl at pH 8.2 and protease sensitive phage were eluted after a variable reaction time. The concentration of protease and incubation times for elution of sensitive phage were decreased gradually over the course of sorting procedure to increase selectivity, with protease concentrations of 0.2 nM (rounds 1-3) and 0.1 nM (rounds 4-9), and reaction times of 5 min (rounds 1-6), 2.5 min (round 7), 40 s (round 8) and 20 s (round 9). Control wells in which no protease was added were also included in each round. For the resistant phage pool, the incubation time with protease remained constant at 5 min. The wells were then washed ten times with PBS plus 0.05% Tween 20 and resistant phage eluted by treatment with 0.1 mL of 0.2 M glycine at pH 2.0 in PBS plus 0.05% Tween*

WO 96/27671

20 for 1 min at room temperature. Protease sensitive and resistant phage pools were titrated and used to infect log phase 27C7 cells for 1 h at 37°C, followed by centrifugation at 4000 rpm, removal of supernatant, and resuspension in 1 mL 2YT medium. The infected cells were then grown 18-24 h in the presence of helper phage as described above and the process repeated 9 times. Selected substrates were introduced into AP fusion proteins and assayed for relative rates of cleavage as described by Matthews and Wells (Matthews, D. J., Goodman, L. J., Gorman, C. M., and Wells, J. A. (1994) *Protein Science* 3:1197-1205 and Matthews, D. J. and Wells, J. A. (1993) *Science* 260:1113-1117), except that the cleavage reactions were performed in 20 mM Tris-Cl at pH 8.2.

#### Example 4

##### 10 Substrate phage selection and cleavage of a fusion protein

Subtilisin has the capability to bind substrates from the P4 to P3' positions (McPhalen, C. A. and James, N. G. (1988) *Biochemistry* 27:6582-6598 and Bode, W., Papamokos, E., Musil, D., Seemueller, U. and Fritz, M. (1986) *EMBO J.* 5:813-818). Given this extensive binding site and the apparent cooperative nature in the way the substrate can bind the enzyme we wished to explore more broadly the substrate preferences for the enzyme. To do this we utilized the substrate phage selection (Matthews, D. J., Goodman, L. J., Gorman, C. M., and Wells, J. A. (1994) *Protein Science* 3:1197-1205 and Matthews, D. J. and Wells, J. A. (1993) *Science* 260:1113-1117) described in Example 3. In this method a five-residue substrate linker that was flanked by di-glycine residues is inserted between an affinity domain (in this case a high affinity variant of hGH) and the carboxy-terminal domain of gene III, a minor coat protein displayed on the surface 15 of the filamentous phage, M13. The five residue substrate linker is fully randomized to generate a library of 20<sup>5</sup> different protein sequence variants. These are displayed on the phage particles which are allowed to bind to the hGHbp. The protease of interest was added and if it cleaved the phage particle at the substrate linker it released that particle. The particles released by protease treatment can be propagated and subjected to another round of selection to further enrich for good protease substrates. Sequences that are retained can 20 also be propagated to enrich for poor protease substrates. By sequencing the isolated phage genes at the end of either selection one can identify good and poor substrates for further analysis.

We chose to focus on the subtilisin BPN' variant N62D/G166D as it was slightly better at discriminating the synthetic dibasic substrates from the others. We subjected the substrate phage library to nine rounds of selection with the subtilisin variant and isolated clones that were either increasingly sensitive 30 or resistant to cleavage. Of twenty-one clones sequenced from the sensitive pool eighteen contained dibasic residues, eleven of which had the substrate linker sequence Asn-Leu-Met-Arg-Lys (SEQ ID NO: 35) (Table 6).

TABLE 6

*Substrate phage sequences sensitive or resistant to N62D/G166D subtilisin from a GG-xxxxx-GG library after 9 rounds of selection<sup>a</sup>.*

## Protease Sensitive Pool

	<u>No Basic Sites (0)</u>	<u>Monobasic Sites (3)</u>	<u>Dibasic Sites (18)</u>
5		N L T A R (3) (SEQ ID NO: 34)	N L M R K (11) (SEQ ID NO: 35)
			T A S R R (6) (SEQ ID NO: 36)
			L T R R S (SEQ ID NO: 37)
			A L S R K (SEQ ID NO: 38)
			L M L R K (SEQ ID NO: 39)

## Protease Resistant Pool

	<u>No Basic Sites (7)</u>	<u>Monobasic Sites (2)</u>	<u>Dibasic Sites (1)</u>
	A S T H F (SEQ ID NO: 40)	Q K P N F (SEQ ID NO: 41)	R K P T H (SEQ ID NO: 42)
10	I Q Q Q Y (SEQ ID NO: 43)	R P G A M (SEQ ID NO: 44)	
	Q G E L P (SEQ ID NO: 47)		
	A P D P T (SEQ ID NO: 46)		
15	Q L L E H (SEQ ID NO: 47)		
	V N N N H (SEQ ID NO: 48)		
20	A Q S N L (SEQ ID NO: 49)		

<sup>a</sup> Numbers in parentheses indicate the number of times a particular DNA sequence was isolated.

Three (3) of the sensitive sequences were monobasic, Asn-Leu-Thr-Ala-Arg (SEQ ID NO: 34). It is known that subtilisin has a preference for hydrophobic residues at the P4 position. If these and the other selected substrates were indeed cleaved after the last basic residue they all would have a Leu, Met or Ala at the P4 position. Almost no basic residues were isolated in the protease resistant pool and those that were had a Pro following the mono- or dibasic residue. It is known that subtilisin does not cleave substrates containing

Pro at the P1' position (Carter, P., Nilsson, B., Burnier, J., Burdick, D. and Wells, J. A. (1989) *Proteins: Struct., Funct., Genet.* 6:240-248). Thus, di-basic substrates were highly selected and these had the additional feature of Leu, Met or Ala at the P4 position.

#### Example 5

##### Cleavage of Substrate Linkers

5

We wished to analyze how efficiently the most frequently selected sequences were cleaved in the context of a fusion protein. For this we applied an alkaline phosphatase-fusion protein assay (Matthews, D. J., Goodman, L. J., Gorman, C. M., and Wells, J. A. (1994) *Protein Science* 3:1197-1205 and Matthews, D. J. and Wells, J. A. (1993) *Science* 260:1113-1117). The hGH substrate linker domains were excised from the phage vector by PCR and fused in front of the gene for *E. coli* AP. The fusion protein was expressed and purified on an hGH receptor affinity column. The fusion protein was bound to the hGH receptor on a plate and treated with the subtilisin variant. The rate of cleavage of the fusion protein from the plate was monitored by collecting soluble fractions as a function of time and assaying for AP activity (Figure 5). The most frequently isolated substrate sequence, Asn-Leu-Met-Arg-Lys (SEQ ID NO: 35) was cleaved about ten times faster than the next most frequently isolated clones (Thr-Ala-Ser-Arg-Arg (SEQ ID NO: 36) and Asn-Leu-Thr-Ala-Arg (SEQ ID NO: 34)). The cleaved AP products were also recovered and subjected to N-terminal sequencing to determine the sites of cleavage (Figure 5), cleavage site denoted by !). In all three fusion proteins, this site was immediately following the dibasic or monobasic site according to the mutant subtilisin design. We also tested the dibasic sequence isolated from the resistant pool, namely Arg-Lys-Pro-Thr-His (SEQ ID NO: 42). We observed no detectable cleavage above background for this substrate during the assay.

The present invention has of necessity been discussed herein by reference to certain specific methods and materials. It is to be understood that the discussion of these specific methods and materials in no way constitutes any limitation on the scope of the present invention, which extends to any and all alternative materials and methods suitable for accomplishing the ends of the present invention.

.....

All references cited herein are expressly incorporated by reference.

## SEQUENCE LISTING

## (1) GENERAL INFORMATION:

(i) APPLICANT: Genentech, Inc.

5 (ii) TITLE OF INVENTION: SUBTILISIN VARIANTS CAPABLE OF CLEAVING  
SUBSTRATES CONTAINING BASIC RESIDUES

(iii) NUMBER OF SEQUENCES: 89

## (iv) CORRESPONDENCE ADDRESS:

(A) ADDRESSEE: Genentech, Inc.  
(B) STREET: 460 Point San Bruno Blvd  
10 (C) CITY: South San Francisco  
(D) STATE: California  
(E) COUNTRY: USA  
(F) ZIP: 94080

## (v) COMPUTER READABLE FORM:

15 (A) MEDIUM TYPE: 3.5 inch, 1.44 Mb floppy disk  
(B) COMPUTER: IBM PC compatible  
(C) OPERATING SYSTEM: PC-DOS/MS-DOS  
(D) SOFTWARE: WinPatin (Genentech)

## (vi) CURRENT APPLICATION DATA:

20 (A) APPLICATION NUMBER:  
(B) FILING DATE:  
(C) CLASSIFICATION:

## (vii) PRIOR APPLICATION DATA:

25 (A) APPLICATION NUMBER: 08/398028  
(B) FILING DATE: 03-MAR-1995

## (viii) ATTORNEY/AGENT INFORMATION:

(A) NAME: Kubinec, Jeffrey S.  
(B) REGISTRATION NUMBER: 36,575  
25 (C) REFERENCE/DOCKET NUMBER: P0936P1PCT

## 30 (ix) TELECOMMUNICATION INFORMATION:

(A) TELEPHONE: 415/225-8228  
(B) TELEFAX: 415/952-9881  
(C) TELEX: 910/371-7168

## (2) INFORMATION FOR SEQ ID NO:1:

## 35 (i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 8119 base pairs  
(B) TYPE: Nucleic Acid  
(C) STRANDEDNESS: Single  
(D) TOPOLOGY: Linear

## 40 (xi) SEQUENCE DESCRIPTION: SEQ ID NO:1:

GAATTCTGGT CTACTAAAAT ATTATTCCAT ACTATACAAT TAATACACAG 50

ATAATCTGT CTATTGGTTA TTCTGCAAAT GAAAAAAAGG AGAGGGATAAA 100

GA GTG AGA GGC AAA AAA GTA TGG ATC AGT TTG CTG TTT 138  
Val Arg Gly Lys Lys Val Trp Ile Ser Leu Leu Phe

45 -107 -105 -100

GCT TTA GCG TTA ATC TTT ACG ATG GCG TTC GGC AGC ACA 177  
Ala Leu Ala Leu Ile Phe Thr Met Ala Phe Gly Ser Thr

	-95	-90	-85
	TCC TCT GCC CAG GCG GCA GGG AAA TCA AAC GGG GAA AAG 216		
	Ser Ser Ala Gln Ala Ala	Gly Lys Ser Asn Gly Glu Lys	
	-80	-75	-70
5	AAA TAT ATT GTC GGG TTT AAA CAG ACA ATG AGC ACG ATG 255		
	Lys Tyr Ile Val Gly Phe Lys Gln Thr Met Ser Thr Met		
	-65	-60	
	AGC GCC GCT AAG AAG AAA GAT GTC ATT TCT GAA AAA GGC 294		
	Ser Ala Ala Lys Lys Lys Asp Val Ile Ser Glu Lys Gly		
10	-55	-50	-45
	GGG AAA GTG CAA AAG CAA TTC AAA TAT GTA GAC GCA GCT 333		
	Gly Lys Val Gln Lys Gln Phe Lys Tyr Val Asp Ala Ala		
	-40	-35	
15	TCA GCT ACA TTA AAC GAA AAA GCT GTA AAA GAA TTG AAA 372		
	Ser Ala Thr Leu Asn Glu Lys Ala Val Lys Glu Leu Lys		
	-30	-25	-20
	AAA GAC CCG AGC GTC GCT TAC GTT GAA GAA GAT CAC GTA 411		
	Lys Asp Pro Ser Val Ala Tyr Val Glu Glu Asp His Val		
	-15	-10	-5
20	GCA CAT GCG TAC GCG CAG TCC GTG CCT TAC GGC GTA TCA 450		
	Ala His Ala Tyr Ala Gln Ser Val Pro Tyr Gly Val Ser		
	1	5	
	CAA ATT AAA GCC CCT GCT CTG CAC TCT CAA GGC TAC ACT 489		
	Gln Ile Lys Ala Pro Ala Leu His Ser Gln Gly Tyr Thr		
25	10	15	20
	GGA TCA AAT GTT AAA GTA GCG GTT ATC GAC AGC GGT ATC 528		
	Gly Ser Asn Val Lys Val Ala Val Ile Asp Ser Gly Ile		
	25	30	35
30	GAT TCT TCT CAT CCT GAT TTA AAG GTA GCA GGC GGA GCC 567		
	Asp Ser Ser His Pro Asp Leu Lys Val Ala Gly Gly Ala		
	40	45	
	AGC ATG GTT CCT TCT GAA ACA AAT CCT TTC CAA GAC AAC 606		
	Ser Met Val Pro Ser Glu Thr Asn Pro Phe Gln Asp Asn		
	50	55	60
35	GAC TCT CAC GGA ACT CAC GTT GCC GGC ACA GTT GCG GCT 645		
	Asp Ser His Gly Thr His Val Ala Gly Thr Val Ala Ala		
	65	70	
	CTT AAT AAC TCA ATC GGT GTA TTA GGC GTT GCG CCA AGC 684		
	Leu Asn Asn Ser Ile Gly Val Leu Gly Val Ala Pro Ser		
40	75	80	85
	GCA TCA CTT TAC GCT GTA AAA GTT CTC GGT GCT GAC GGT 723		
	Ala Ser Leu Tyr Ala Val Lys Val Leu Gly Ala Asp Gly		
	90	95	100
45	TCC GGC CAA TAC AGC TGG ATC ATT AAC GGA ATC GAG TGG 762		
	Ser Gly Gln Tyr Ser Trp Ile Ile Asn Gly Ile Glu Trp		
	105	110	
	GCG ATC GCA AAC AAT ATG GAC GTT ATT AAC ATG AGC CTC 801		
	Ala Ile Ala Asn Asn Met Asp Val Ile Asn Met Ser Leu		
	115	120	125

GGC GGA CCT TCT GGT TCT GCT GCT TTA AAA GCG GCA GTT 840  
 Gly Gly Pro Ser Gly Ser Ala Ala Leu Lys Ala Ala Val  
 130 135

5 GAT AAA GCC GTT GCA TCC GGC GTC GTA GTC GTT GCG GCA 879  
 Asp Lys Ala Val Ala Ser Gly Val Val Val Val Ala Ala  
 140 145 150

GCC GGT AAC GAA GGC ACT TCC GGC AGC TCG TCG ACA GTG 918  
 Ala Gly Asn Glu Gly Thr Ser Gly Ser Ser Ser Thr Val  
 155 160 165

10 GAC TAC CCT GGC AAA TAC CCT TCT GTC ATT GCA GTA GGC 957  
 Asp Tyr Pro Gly Lys Tyr Pro Ser Val Ile Ala Val Gly  
 170 175

GCT GTT GAC AGC AGC AAC CAA AGA GCA TCT TTC TCA AGC 996  
 Ala Val Asp Ser Ser Asn Gin Arg Ala Ser Phe Ser Ser  
 15 180 185 190

GTA GGA CCT GAG CTT GAT GTC ATG GCA CCT GGC GTA TCT 1035  
 Val Gly Pro Glu Leu Asp Val Met Ala Pro Gly Val Ser  
 195 200

20 ATC CAA AGC ACG CTT CCT GGA AAC AAA TAC GGG GCG TAC 1074  
 Ile Gln Ser Thr Leu Pro Gly Asn Lys Tyr Gly Ala Tyr  
 205 210 215

AAC GGT ACC TCA ATG GCA TCT CCG CAC GTT GCC GGA GCG 1113  
 Asn Gly Thr Ser Met Ala Ser Pro His Val Ala Gly Ala  
 220 225 230

25 GCT GCT TTG ATT CTT TCT AAG CAC CCG AAC TGG ACA AAC 1152  
 Ala Ala Leu Ile Leu Ser Lys His Pro Asn Trp Thr Asn  
 235 240

ACT CAA GTC CGC AGC AGT TTA GAA AAC ACC ACT ACA AAA 1191  
 Thr Gln Val Arg Ser Ser Leu Glu Asn Thr Thr Lys  
 30 245 250 255

CTT GGT GAT TCT TTC TAC TAT GGA AAA GGG CTG ATC AAC 1230  
 Leu Gly Asp Ser Phe Tyr Tyr Gly Lys Gly Leu Ile Asn  
 260 265

35 GTA CAG GCG GCA GCT CAG TA AAACATAAAA AACCGGCCCTT 1270  
 Val Gln Ala Ala Ala Gln  
 270 275

GGCCCCGCCG GTTTTTATT ATTTTTCTTC CTCCGCATGT TCAATCCGCT 1320  
 CCATAATCGA CGGATGGCTC CCTCTGAAAA TTTAACGAG AACGGCGGG 1370  
 TTGACCCGGC TCAGTCCCCT AACGGCCAAG TCCTGAAACG TCTCAATCGC 1420

40 CGCTTCCGG TTTCCGGTCA GCTCAATGCC GTAACGGCTCG GCGGCCTTT 1470  
 CCTGATAACCG GGAGACGGCA TTCGTAATCG GATCCGGAAA TTGTAAACGT 1520  
 TAATATTTG TTAATTTCG CGTTAAATT TTGTTAAATC AGCTCATTTT 1570  
 TTAACCAATA GGCGAAATC GGCAAAATCC CTTATAAATC AAAAGAATAG 1620  
 ACCGAGATAG GGTTGAGTGT TGTTCCAGTT TGGAACAGA GTCCACTATT 1670

45 AAAGAACGTG GACTCCAACG TCAAAGGGCG AAAACCGTC TATCAGGGCT 1720

WO 96/27671

ATGGCCCACT ACGTGAACCA TCACCCATAAT CAAGTTTTT GGGGTCGAGG 1770  
 TGCCGTAAAG CACTAAATCG GAACCTAAA GGGAGCCCCC GATTTAGAGC 1820  
 TTGACGGGGA AAGCCGGCGA ACGTGGCGAG AAAGGAAGGG AAGAAAGCGA 1870  
 AAGGAGCGGG CGCTAGGGCG CTGGCAAGTG TAGCGGTAC GCTGCGCGTA 1920  
 5 ACCACCCACAC CCGCCGCGCT TAATGCGCCG CTACAGGGCG CGTCCGGATC 1970  
 NGATCCGACG CGAGGCTGGA TGGCCTCCC CATTATGATT CTTCTCGCTT 2020  
 CCGGCGGCAT CGGGATSCCC GCGTTGCAGG CCATGCTGTC CAGGCAGGTA 2070  
 GATGACGACC ATCAGGGACA GCTTCAAGGA TCGCTCGCG CTCTTACCAAG 2120  
 CCTAACCTCG ATCACTGGAC CGCTGATCGT CACGGCGATT TATGCCGCCT 2170  
 10 CGGGCAGCAC ATGGAACGGG TTGGCATGGA TTGTAGGCAG CGCCCTATAC 2220  
 CTTGTCTGCC TCCCCGCGTT GCGTCGCGGT GCATGGAGCC GGGCCACCTC 2270  
 GACCTGAATG GAAGCCGGCG GCACCTCGCT AACGGATTCA CCACTCCAAG 2320  
 AATTGGAGCC AATCAATTCT TGC GGAGAAC TGTGAATGCG CAAACCAACC 2370  
 CTTGGCAGAA CATATCCATC GCGTCCGCCA TCTCCAGCAG CCGCACGGCG 2420  
 15 CGCATCTCGG GCCGCGTTGC TGGCGTTTT CCATAGGCTC CGCCCCCCTG 2470  
 ACGAGCATCA CAAAAATCGA CGCTCAAGTC AGAGGTGGCG AAACCCGACA 2520  
 GGACTATAAA GATACCAGGC GTTCCCCCT GGAAGCTCCC TCGTGCGCTC 2570  
 TCCTGTTCCG ACCCTGCCGC TTACCGGATA CCTGTCCGCC TTTCTCCCTT 2620  
 CGGGAAAGCGT GGCGCTTCT CAATGCTCAC GCTGTAGGTA TCTCAGTTCG 2670  
 20 GTGTAGGTCTG TTCGCTCCAA GCTGGGCTGT GTGCACGAAC CCCCCGTTCA 2720  
 GCCCGACCGC TGCGCCTTAT CCGGTAACTA TCGTCTTGAG TCCAACCCGG 2770  
 TAAGACACGA CTTATGCCA CTGGCAGCAG CCACTGGTAA CAGGATTAGC 2820  
 AGAGCGAGGT ATGTAGGGGG TGCTACAGAG TTCTTGAAGT GGTGGCTAA 2870  
 CTACGGCTAC ACTAGAAGGA CAGTATTGG TATCTCGCT CTGCTGAAGC 2920  
 25 CAGTTACCTT CGGAAAAAGA GTTGGTAGCT CTTGATCCGG CAAACAAACC 2970  
 ACCGCTGGTA GCGGTGGTTT TTTTGTTCG AAGCAGCAGA TTACGCGCAG 3020  
 AAAAAAAAGGA TCTCAAGAAG ATCCTTTGAT CTTTCTACG GGGTCTGACG 3070  
 CTCAGTGGAA CGAAAACCTCA CGTTAAGGGAA TTTTGGTCAT GAGATTATCA 3120  
 AAAAGGATCT TCACCTAGAT CCTTTAAAT TAAAAATGAA GTTTAAATC 3170  
 30 AATCTAAAGT ATATATGAGT AAACCTGGTC TGACAGTTAC CAATGCTTAA 3220  
 TCAGTGAGGC ACCTATCTCA GCGATCTGTC TATTCGTTTC ATCCATAGTT 3270  
 GCCTGACTCC CCGTCGTGTA GATAACTACG ATACGGGAGG GCTTACCATC 3320  
 TGGCCCCAGT GCTGCAATGA TACCGCGAGA CCCACGCTCA CCGGCTCCAG 3370

ATTTATCAGC AATAAACAG CCAGCCGAA GGGCGAGCG CAGAAGTGGT 3420  
CCTGCAACTT TATCCGCCCTC CATCCAGTGT ATTAATTGTT GCCGGGAAGC 3470  
TAGAGTAAGT AGTTCCCCAG TTAAATAGTT GCCCAACCTT GTGCCATTG 3520  
CTGCAGGCAT CGTGGTGTCA CGCTCGTCGT TTGGTATGGC TTCATTCAAGC 3570  
5 TCCGGTTCCC AACGATCAAG GCGAGTTACA TGATCCCCA TGGTGTGCAA 3620  
AAAAGCGGTT AGCTCCTTCG GTCCCTCCGAT CGTGTGAGA AGTAAGTGG 3670  
CCGCAGTGTT ATCACTCATG GTTATGGCAG CACTGCATAA TTCTCTTACT 3720  
GTCATGCCAT CGCTAAGATG CTTTCTGTG ACTGGTGAGT ACTCAACCAA 3770  
GTCATTCTGA GAATAGTGTG TGCGGCGACC GAGTTGCTCT TGCCCGCGT 3820  
10 CAACACGGGA TAATACCGC CCACATAGCA GAACTTAAA AGTGCTCATC 3870  
ATGGAAAAC GTTCTTCGGG GCGAAAACTC TCAAGGATCT TACCGCTGTT 3920  
GAGATCCAGT TCGATGTAAC CCACTCGTGC ACCCAACTGA TCTTCAGCAT 3970  
CTTTTACTTT CACCAGCGTT TCTGGGTGAG CAAAAACAGG AAGGCAAAAT 4020  
GCCGAAAAAA AGGGAATAAG GGCGACACGG AAATGTTGAA TACTCATACT 4070  
15 CTTCCCTTTT CAATATTATT GAAGCATTAA TCAGGGTTAT TGTCTCATGA 4120  
GCGGATACAT ATTTGAATGT ATTTAGAAAA ATAACAAAT AGGGGTTCCG 4170  
CCGACATTTC CCCGAAAAGT GCCACCTGAC GTCTAAGAAA CCATTATTAT 4220  
CATGACATTA ACCTATAAAA ATAGGCATAT CACGAGGCC CTTCGTCTTC 4270  
AAGAATTAAAT CCTTTAAGGA ACGTACAGAC GGCTTAAAG CCTTTAAAAA 4320  
20 CGTTTTAAG GGGTTGTAG ACAAGGTTAA GGATAAAACA GCACAATTCC 4370  
AAGAAAAACA CGATTTAGAA CCTAAAAAGA ACGAATTGAA ACTAACTCAT 4420  
AACCGAGAGG TAAAAAAAGA ACGAAGTCGA GATCAGGGAA TGAGTTATA 4470  
AAATAAAAAA AGCACCTGAA AAGGTGTCTT TTTTGATGG TTTGAACTT 4520  
GTTCTTCTT ATCTTGATAC ATATAGAAAT AACGTCTTT TTATTTAGT 4570  
25 TGCTGAAAGG TCGTTGAAG TGTTGGTATG TATGTGTTT AAAGTATTGA 4620  
AAACCCCTAA AATTGGTGC ACAGAAAAAC CCCATCTGTT AAAGTTATAA 4670  
GTGACTAAAC AAATAACTAA ATAGATGGG GTTCTTTA ATATTATGT 4720  
TCCTAATAGT AGCATTTATT CAGATGAAA ATCAAGGTT TTAGTGGACA 4770  
AGACAAAAAG TGGAAAAGTG AGACCATGGA GAGAAAAGAA AATCGCTAAT 4820  
30 GTTGATTACT TTGAACCTCT GCATATTCTT GAATTTAAA AGGCTGAAAG 4870  
AGTAAAAAGAT TGTGCTGAAA TATTAGAGTA TAAACAAAT CGTGAACAG 4920  
GCGAAAAGAAA GTTGTATCGA GTGTGGTTT GTAAATCCAG GCTTTGTCCA 4970  
ATGTGCAACT GGAGGAGAGC AATGAAACAT GGCATTCACT CACAAAAGGT 5020

WO 96/27671

TGTTGCTGAA GTTATTAAAC AAAAGCCAAC AGTCGTTGG TTGTTTCTCA 5070  
 CATTAAACAGT TAAAAATGTT TATGATGGCG AAGAATTAAA TAAGAGTTG 5120  
 TCAGATATGG CTCAGGATT TCGCCGAATG ATGCAATATA AAAAAATTAA 5170  
 TAAAAATCTT GTGGGTTTA TGCGTGCAAC GGAAGTGACA ATAAATAATA 5220  
 5 AAGATAATTC TTATAATCAG CACATGCATG TATTGGTATG TGTGGAACCA 5270  
 ACTTATTTA AGAATACAGA AACTACGTG AATCAAAAAC AATGGATTCA 5320  
 ATTTGGAAA AAGGCAATGA AATTAGACTA TGATCCAAAT CTAAAAGTTC 5370  
 AAATGATTG ACCGAAAAAT AAATATAAAAT CGGATATACA ATCGGCAATT 5420  
 GACGAAACTG CAAAATATCC TGAAAGGAT ACGGATTTA TGACCGATGA 5470  
 10 TGAAAGAAAAG AATTGAAAC GTTGTCTGA TTTGGAGGAA GGTTTACACC 5520  
 GTAAAAGGTT AATCTCCTAT GGTGGTTGT TAAAAGAAAAT ACATAAAAAA 5570  
 TAAACCTTG ATGACACAGA AGAAGGCGAT TTGATTCTA CAGATGATGA 5620  
 CGAAAAAGCC GATGAAGATG GATTTCTAT TATTGCAATG TGGAATTGGG 5670  
 AACGGAAAAA TTATTTATT AAAGAGTAGT TCAACAAACG GGCCAGTTG 5720  
 15 TTGAAGATTA GATGCTATAA TTGTTATTAA AAGGATTGAA GGATGCTTAG 5770  
 GAAGACGAGT TATTAATAGC TGAATAAGAA CGGTGCTCTC CAAATATTCT 5820  
 TATTAGAAA AGCAAATCTA AAATTATCTG AAAAGGGAAT GAGAATAGTG 5870  
 AATGGACAA TAATAATGAC TAGAGAAGAA AGAATGAAGA TTGTTCATGA 5920  
 AATTAAGGAA CGAATATTGG ATAAATATGG GGATGATGTT AAGGCTATTG 5970  
 20 GTGTTATGG CTCTCTTGGT CGTCAGACTG ATGGGCCCTA TTCGGATATT 6020  
 GAGATGATGT GTGTCATGTC AACAGAGGAA GCAGAGTTCA GCCATGAATG 6070  
 GACAACCGGT GAGTGGAGG TGGAAGTGAA TTTGATAGC GAAGAGATT 6120  
 TACTAGATTA TGCATCTAG GTGGAATCAG ATTGGCCGCT TACACATGGT 6170  
 CAATTTTCT CTATTTGCC GATTTATGAT TCAGGTGGAT ACTTAGAGAA 6220  
 25 AGTGTATCAA ACTGCTAAAT CGGTAGAAC CCAAACGTTC CACGATGCGA 6270  
 TTTGTGCCCT TATCGTAGAA GAGCTGTTG AATATGCAGG CAAATGGCGT 6320  
 AATATTCGTG TGCAAGGACC GACAACATTT CTACCATCT TGACTGTACA 6370  
 GGTAGCAATG GCAGGTGCCA TGTTGATTGG TCTGCATCAT CGCATCTGTT 6420  
 ATACGACGAG CGCTTCGGTC TTAACGTAGA CAGTTAAGCA ATCAGATCTT 6470  
 30 CCTTCAGGTT ATGACCATCT GTGCCAGTTC GTAATGTCTG GTCAACTTTC 6520  
 CGACTCTGAG AAACTTCTGG AATCGCTAGA GAATTTCTGG AATGGGATTG 6570  
 AGGAGTGGAC AGAACGACAC GGATATATAG TGGATGTGTC AAAACGCCATA 6620  
 CCATTTGAA CGATGACCTC TAATAATTGT TAATCATGTT GGTTACGTAT 6670

TTATTAACCT CTCCTAGTAT TAGTAATTAT CATGGCTGTC ATGGCCGATT 6720  
AACGGAATAA AGGGTGTGCT TAAATCGGGC CATTTCGCGT AATAAGAAAA 6770  
AGGATTAATT ATGAGCGAAT TGAATTAATA ATAAGGTAAT AGATTACAT 6820  
TAGAAAATGA AAGGGGATTT TATGCGTGAG AATGTTACAG TCTATCCCG 6870  
5 CAATAGTTAC CCTTATTATC AAGATAAGAA AGAAAAGGAT TTTTCGCTAC 6920  
GCTCAAATCC TTTAAAAAAA CACAAAAGAC CACATTTTT AATGTTGGTCT 6970  
TTATTCTTCA ACTAAAGCAC CCATTAGTTC AACAAACGAA AATTGGATAA 7020  
AGTGGGATAT TTTTAAAATA TATATTTATG TTACAGTAAT ATTGACTTT 7070  
AAAAAAAGGAT TGATTCTAAT GAAGAAAGCA GACAAGTAAG CCTCCCTAAAT 7120  
10 TCACTTTAGA TAAAATTTA GGAGGCATAT CAAATGAAC TTAATAAAAT 7170  
TGATTTAGAC AATTGGAAGA GAAAAGAGAT ATTTAATCAT TATTGAAACC 7220  
AACAAACGAC TTTTAGTATA ACCACAGAAA TTGATATTAG TGTTTATAC 7270  
CGAAACATAA AACAAAGAAGG ATATAAATTT TACCCCTGCAT TTATTTCTT 7320  
AGTGACAAGG GTGATAAACT CAAATACAGC TTTAGAACT GGTTACAATA 7370  
15 GCGACGGAGA GTTAGGTTAT TGGGATAAGT TAGAGCCACT TTATACAATT 7420  
TTTGATGGTG TATCTAAAC ATTCTCTGGT ATTTGGACTC CTGTAAAGAA 7470  
TGACTTCAAA GAGTTTTATG ATTTATACCT TTCTGATGTA GAGAAATATA 7520  
ATGGTTCGGG GAAATTGTTT CCCAAACAC CTATACCTGA AAATGCTTT 7570  
TCTCTTCTA TTATTCCATG GACTTCATTI ACTGGGTTTA ACTTAAATAT 7620  
20 CAATAATAAT AGTAATTACC TTCTACCCAT TATTACAGCA GGAAAATTCA 7670  
TTAATAAAAGG TAATTCAATA TATTTACCGC TATCTTACA GGTACATCAT 7720  
TCTGTTGTG ATGGTTATCA TGCAAGGATTG TTTATGAACT CTATTCAGGA 7770  
ATTGTCAGAT AGGCCATAATG ACTGGCTTT ATAATATGAG ATAATGCCGA 7820  
CTGTACTTTT TACAGTCGGT TTTCTAATGT CACTAACCTG CCCCCTTAGT 7870  
25 TGAAGAAGGT TTTTATATTA CAGCTCCAGA TCCATATCCT TCTTTTCTG 7920  
AACCGACTTC TCCTTTTCCG CTTCTTTATT CCAATTGCTT TATTGACGTT 7970  
GAGCCTCGGA ACCCNATAG TGTGTTATAC TTTACTTGGA AGTGGTTGCC 8020  
GGAAAGAGCG AAAATGCCTC ACATTGTGC CACCTAAAAA GGAGCGATT 8070  
ACATATGAGT TATGCAAGTT GTAGAATGCA AAAAGTGAAGA TCAGGATCN 8119  
30 (2) INFORMATION FOR SEQ ID NO:2:  
  
(i) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 382 amino acids  
(B) TYPE: Amino Acid  
(D) TOPOLOGY: Linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:2:

Val Arg Gly Lys Lys Val Trp Ile Ser Leu Leu Phe Ala Leu Ala  
 -107 -105 -100 -95  
 Leu Ile Phe Thr Met Ala Phe Gly Ser Thr Ser Ser Ala Gln Ala  
 5 -90 -85 -80  
 Ala Gly Lys Ser Asn Gly Glu Lys Lys Tyr Ile Val Gly Phe Lys  
 -75 -70 -65  
 Gln Thr Met Ser Thr Met Ser Ala Ala Lys Lys Lys Asp Val Ile  
 -60 -55 -50  
 10 Ser Glu Lys Gly Lys Val Gln Lys Gln Phe Lys Tyr Val Asp  
 -45 -40 -35  
 Ala Ala Ser Ala Thr Leu Asn Glu Lys Ala Val Lys Glu Leu Lys  
 -30 -25 -20  
 Lys Asp Pro Ser Val Ala Tyr Val Glu Glu Asp His Val Ala His  
 15 -15 -10 -5  
 Ala Tyr Ala Gln Ser Val Pro Tyr Gly Val Ser Gln Ile Lys Ala  
 1 5 10  
 Pro Ala Leu His Ser Gln Gly Tyr Thr Gly Ser Asn Val Lys Val  
 15 20 25  
 20 Ala Val Ile Asp Ser Gly Ile Asp Ser Ser His Pro Asp Leu Lys  
 30 35 40  
 Val Ala Gly Gly Ala Ser Met Val Pro Ser Glu Thr Asn Pro Phe  
 45 50 55  
 Gln Asp Asn Asp Ser His Gly Thr His Val Ala Gly Thr Val Ala  
 25 60 65 70  
 Ala Leu Asn Asn Ser Ile Gly Val Leu Gly Val Ala Pro Ser Ala  
 75 80 85  
 Ser Leu Tyr Ala Val Lys Val Leu Gly Ala Asp Gly Ser Gly Gln  
 90 95 100  
 30 Tyr Ser Trp Ile Ile Asn Gly Ile Glu Trp Ala Ile Ala Asn Asn  
 105 110 115  
 Met Asp Val Ile Asn Met Ser Leu Gly Gly Pro Ser Gly Ser Ala  
 120 125 130  
 35 Ala Leu Lys Ala Ala Val Asp Lys Ala Val Ala Ser Gly Val Val  
 135 140 145  
 Val Val Ala Ala Ala Gly Asn Glu Gly Thr Ser Gly Ser Ser Ser  
 150 155 160  
 Thr Val Asp Tyr Pro Gly Lys Tyr Pro Ser Val Ile Ala Val Gly  
 165 170 175  
 40 Ala Val Asp Ser Ser Asn Gln Arg Ala Ser Phe Ser Ser Val Gly  
 180 185 190  
 Pro Glu Leu Asp Val Met Ala Pro Gly Val Ser Ile Gln Ser Thr  
 195 200 205

Leu Pro Gly Asn Lys Tyr Gly Ala Tyr Asn Gly Thr Ser Met Ala  
210 215 220

Ser Pro His Val Ala Gly Ala Ala Leu Ile Leu Ser Lys His  
225 230 235

5 Pro Asn Trp Thr Asn Thr Gln Val Arg Ser Ser Leu Glu Asn Thr  
240 245 250

Thr Thr Lys Leu Gly Asp Ser Phe Tyr Tyr Gly Lys Gly Leu Ile  
255 260 265

10 Asn Val Gln Ala Ala Ala Gln  
270 275

(2) INFORMATION FOR SEQ ID NO:3:

(i) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 7 amino acids  
(B) TYPE: Amino Acid  
15 (D) TOPOLOGY: Linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:3:

Ser Leu Gly Gly Pro Ser Gly  
1 5 7

(2) INFORMATION FOR SEQ ID NO:4:

20 (i) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 7 amino acids  
(B) TYPE: Amino Acid  
(D) TOPOLOGY: Linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:4:

25 Ala Ala Ala Gly Asn Glu Gly  
1 5 7

(2) INFORMATION FOR SEQ ID NO:5:

30 (i) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 6 amino acids  
(B) TYPE: Amino Acid  
(D) TOPOLOGY: Linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:5:

Ser Thr Val Gly Tyr Pro  
1 5 6

35 (2) INFORMATION FOR SEQ ID NO:6:

(i) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 7 amino acids  
(B) TYPE: Amino Acid  
(D) TOPOLOGY: Linear

40 (xi) SEQUENCE DESCRIPTION: SEQ ID NO:6:

Ser Trp Gly Pro Ala Asp Asp  
1 5 7

(2) INFORMATION FOR SEQ ID NO:7:

(i) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 7 amino acids  
(B) TYPE: Amino Acid  
(D) TOPOLOGY: Linear

5 (xi) SEQUENCE DESCRIPTION: SEQ ID NO:7:

Phe Ala Ser Gly Asn Gly Gly  
1 5 7

(2) INFORMATION FOR SEQ ID NO:8:

10 (i) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 7 amino acids  
(B) TYPE: Amino Acid  
(D) TOPOLOGY: Linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:8:

15 Cys Asn Tyr Asp Gly Tyr Thr  
1 5 7

(2) INFORMATION FOR SEQ ID NO:9:

20 (i) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 7 amino acids  
(B) TYPE: Amino Acid  
(D) TOPOLOGY: Linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:9:

Ser Trp Gly Pro Glu Asp Asp  
1 5 7

(2) INFORMATION FOR SEQ ID NO:10:

25 (i) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 7 amino acids  
(B) TYPE: Amino Acid  
(D) TOPOLOGY: Linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:10:

30 Trp Ala Ser Gly Asn Gly Gly  
1 5 7

(2) INFORMATION FOR SEQ ID NO:11:

35 (i) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 7 amino acids  
(B) TYPE: Amino Acid  
(D) TOPOLOGY: Linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:11:

Cys Asn Cys Asp Gly Tyr Thr  
1 5 7

40 (2) INFORMATION FOR SEQ ID NO:12:

(i) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 7 amino acids  
(B) TYPE: Amino Acid  
(D) TOPOLOGY: Linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:12:

Trp Ala Ser Gly Asp Gly Gly  
1 5 7

(2) INFORMATION FOR SEQ ID NO:13:

5 (i) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 7 amino acids  
(B) TYPE: Amino Acid  
(D) TOPOLOGY: Linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:13:

10 Cys Asn Cys Asp Gly Tyr Ala  
1 5 7

(2) INFORMATION FOR SEQ ID NO:14:

15 (i) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 6 amino acids  
(B) TYPE: Amino Acid  
(D) TOPOLOGY: Linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:14:

Val Ile Asp Ser Gly Ile  
1 5 6

20 (2) INFORMATION FOR SEQ ID NO:15:

(i) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 5 amino acids  
(B) TYPE: Amino Acid  
(D) TOPOLOGY: Linear

25 (xi) SEQUENCE DESCRIPTION: SEQ ID NO:15:

Asp Asn Asn Ser His  
1 5

(2) INFORMATION FOR SEQ ID NO:16:

30 (i) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 6 amino acids  
(B) TYPE: Amino Acid  
(D) TOPOLOGY: Linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:16:

35 Ile Val Asp Asp Gly Leu  
1 5 6

(2) INFORMATION FOR SEQ ID NO:17:

(i) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 5 amino acids  
(B) TYPE: Amino Acid  
(D) TOPOLOGY: Linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:17:

Ser Asp Asp Tyr His  
1 5

## (2) INFORMATION FOR SEQ ID NO:18:

5 (i) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 6 amino acids  
(B) TYPE: Amino Acid  
(D) TOPOLOGY: Linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:18:

Ile Leu Asp Asp Gly Ile  
1 5 6

## (2) INFORMATION FOR SEQ ID NO:19:

10 (i) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 5 amino acids  
(B) TYPE: Amino Acid  
(D) TOPOLOGY: Linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:19:

15 Asn Asp Asn Arg His  
1 5

## (2) INFORMATION FOR SEQ ID NO:20:

20 (i) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 6 amino acids  
(B) TYPE: Amino Acid  
(D) TOPOLOGY: Linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:20:

Ile Met Asp Asp Gly Ile  
1 5 6

## 25 (2) INFORMATION FOR SEQ ID NO:21:

(i) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 5 amino acids  
(B) TYPE: Amino Acid  
(D) TOPOLOGY: Linear

30 (xi) SEQUENCE DESCRIPTION: SEQ ID NO:21:

Trp Phe Asn Ser His  
1 5

## (2) INFORMATION FOR SEQ ID NO:22:

35 (i) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 27 base pairs  
(B) TYPE: Nucleic Acid  
(C) STRANDEDNESS: Single  
(D) TOPOLOGY: Linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:22:

40 GCGGTTATCG ACGACGGTAT CGATTCT 27

## (2) INFORMATION FOR SEQ ID NO:23:

(i) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 27 base pairs

WO 96/27671

- (B) TYPE: Nucleic Acid
- (C) STRANDEDNESS: Single
- (D) TOPOLOGY: Linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:23:

5 GCGGTTATCG ACAAAAGGTAT CGATTCT 27

(2) INFORMATION FOR SEQ ID NO:24:

- (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 27 base pairs
- (B) TYPE: Nucleic Acid
- (C) STRANDEDNESS: Single
- (D) TOPOLOGY: Linear

10 (xi) SEQUENCE DESCRIPTION: SEQ ID NO:24:

GCGGTTATCG ACGAAGGTAT CGATTCT 27

(2) INFORMATION FOR SEQ ID NO:25:

- (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 23 base pairs
- (B) TYPE: Nucleic Acid
- (C) STRANDEDNESS: Single
- (D) TOPOLOGY: Linear

15 20 (xi) SEQUENCE DESCRIPTION: SEQ ID NO:25:

CCAAGACAAAC GACTCTCACCG GAA 23

(2) INFORMATION FOR SEQ ID NO:26:

- (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 23 base pairs
- (B) TYPE: Nucleic Acid
- (C) STRANDEDNESS: Single
- (D) TOPOLOGY: Linear

25 (xi) SEQUENCE DESCRIPTION: SEQ ID NO:26:

CCAAGACAAAC AGCTCTCACCG GAA 23

30 (2) INFORMATION FOR SEQ ID NO:27:

- (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 23 base pairs
- (B) TYPE: Nucleic Acid
- (C) STRANDEDNESS: Single
- (D) TOPOLOGY: Linear

35 (xi) SEQUENCE DESCRIPTION: SEQ ID NO:27:

CCAAGACAAAC AAATCTCACCG GAA 23

(2) INFORMATION FOR SEQ ID NO:28:

- (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 42 base pairs

WO 96/27671

- (B) TYPE: Nucleic Acid
- (C) STRANDEDNESS: Single
- (D) TOPOLOGY: Linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:28:

5 CACTTCCGGC AGCTCGTCGA CAGTGGACTA CCCTGGAAA TA 42

(2) INFORMATION FOR SEQ ID NO:29:

- (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 42 base pairs
- (B) TYPE: Nucleic Acid
- (C) STRANDEDNESS: Single
- (D) TOPOLOGY: Linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:29:

CACTTCCGGC AGCTCGTCGA CAGTGGAGTA CCCTGGAAA TA 42

(2) INFORMATION FOR SEQ ID NO:30:

15 (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 41 base pairs
- (B) TYPE: Nucleic Acid
- (C) STRANDEDNESS: Single
- (D) TOPOLOGY: Linear

20 (xi) SEQUENCE DESCRIPTION: SEQ ID NO:30:

TTAACATGAG CCTCGGCCCA GCTAGCGGTT CTGCTGCTTT A 41

(2) INFORMATION FOR SEQ ID NO:31:

- (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 43 base pairs
- (B) TYPE: Nucleic Acid
- (C) STRANDEDNESS: Single
- (D) TOPOLOGY: Linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:31:

TTAACATGAG CCTCGGCCCA GCGGATGATT CTGCTGCTTT AAA 43

30 (2) INFORMATION FOR SEQ ID NO:32:

- (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 47 base pairs
- (B) TYPE: Nucleic Acid
- (C) STRANDEDNESS: Single
- (D) TOPOLOGY: Linear

35 (xi) SEQUENCE DESCRIPTION: SEQ ID NO:32:

CGGCAGCTCA AGCAACGATG GCTATCCTGG CAAATACCT TCTGTCA 47

(2) INFORMATION FOR SEQ ID NO:33:

40 (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 44 base pairs

- (B) TYPE: Nucleic Acid
- (C) STRANDEDNESS: Single
- (D) TOPOLOGY: Linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:33:

5       ACTTCCGGCA GCTCTTCGAA CTACGACGGG TACCCTGGCA AATA 44

(2) INFORMATION FOR SEQ ID NO:34:

- (i) SEQUENCE CHARACTERISTICS:
  - (A) LENGTH: 5 amino acids
  - (B) TYPE: Amino Acid
  - (D) TOPOLOGY: Linear

10

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:34:

Asn Leu Thr Ala Arg  
1                         5

(2) INFORMATION FOR SEQ ID NO:35:

15

- (i) SEQUENCE CHARACTERISTICS:
  - (A) LENGTH: 5 amino acids
  - (B) TYPE: Amino Acid
  - (D) TOPOLOGY: Linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:35:

20

Asn Leu Met Arg Lys  
1                         5

(2) INFORMATION FOR SEQ ID NO:36:

25

- (i) SEQUENCE CHARACTERISTICS:
  - (A) LENGTH: 5 amino acids
  - (B) TYPE: Amino Acid
  - (D) TOPOLOGY: Linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:36:

Thr Ala Ser Arg Arg  
1                         5

30

(2) INFORMATION FOR SEQ ID NO:37:

- (i) SEQUENCE CHARACTERISTICS:
  - (A) LENGTH: 5 amino acids
  - (B) TYPE: Amino Acid
  - (D) TOPOLOGY: Linear

35

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:37:

Leu Thr Arg Arg Ser  
1                         5

(2) INFORMATION FOR SEQ ID NO:38:

40

- (i) SEQUENCE CHARACTERISTICS:
  - (A) LENGTH: 5 amino acids
  - (B) TYPE: Amino Acid
  - (D) TOPOLOGY: Linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:38:

WO 96/27671

Ala Leu Ser Arg Lys  
 1                       5

## (2) INFORMATION FOR SEQ ID NO:39:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 5 amino acids
- (B) TYPE: Amino Acid
- (D) TOPOLOGY: Linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:39:

Leu Met Leu Arg Lys  
 10                  1                       5

## (2) INFORMATION FOR SEQ ID NO:40:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 5 amino acids
- (B) TYPE: Amino Acid
- (D) TOPOLOGY: Linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:40:

Ala Ser Thr His Phe  
 1                       5

## (2) INFORMATION FOR SEQ ID NO:41:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 5 amino acids
- (B) TYPE: Amino Acid
- (D) TOPOLOGY: Linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:41:

Gln Lys Pro Asn Phe  
 25                  1                       5

## (2) INFORMATION FOR SEQ ID NO:42:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 5 amino acids
- (B) TYPE: Amino Acid
- (D) TOPOLOGY: Linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:42:

Arg Lys Pro Thr His  
 30                  1                       5

## 35 (2) INFORMATION FOR SEQ ID NO:43:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 5 amino acids
- (B) TYPE: Amino Acid
- (D) TOPOLOGY: Linear

## 40 (xi) SEQUENCE DESCRIPTION: SEQ ID NO:43:

Ile Gln Gln Gln Tyr  
 1                       5

## (2) INFORMATION FOR SEQ ID NO:44:

WO 96/27671

(i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 5 amino acids  
 (B) TYPE: Amino Acid  
 (D) TOPOLOGY: Linear

5 (xi) SEQUENCE DESCRIPTION: SEQ ID NO:44:

Arg Pro Gly Ala Met  
 1 5

10 (2) INFORMATION FOR SEQ ID NO:45:

(i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 5 amino acids  
 (B) TYPE: Amino Acid  
 (D) TOPOLOGY: Linear

15 (xi) SEQUENCE DESCRIPTION: SEQ ID NO:45:

Gln Gly Glu Leu Pro  
 1 5

20 (2) INFORMATION FOR SEQ ID NO:46:

(i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 5 amino acids  
 (B) TYPE: Amino Acid  
 (D) TOPOLOGY: Linear

25 (xi) SEQUENCE DESCRIPTION: SEQ ID NO:46:

Ala Pro Asp Pro Thr  
 1 5

30 (2) INFORMATION FOR SEQ ID NO:47:

25 (i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 5 amino acids  
 (B) TYPE: Amino Acid  
 (D) TOPOLOGY: Linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:47:

35 Gln Leu Leu Glu His  
 1 5

40 (2) INFORMATION FOR SEQ ID NO:48:

(i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 5 amino acids  
 (B) TYPE: Amino Acid  
 (D) TOPOLOGY: Linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:48:

45 Val Asn Asn Asn His  
 1 5

50 (2) INFORMATION FOR SEQ ID NO:49:

(i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 5 amino acids  
 (B) TYPE: Amino Acid  
 (D) TOPOLOGY: Linear

WO 96/27671

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:49:

Ala Gln Ser Asn Leu  
1 5

(2) INFORMATION FOR SEQ ID NO:50:

5 (i) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 5 amino acids  
(B) TYPE: Amino Acid  
(D) TOPOLOGY: Linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:50:

10 Thr Ala Ser Arg Arg  
1 5

(2) INFORMATION FOR SEQ ID NO:51:

15 (i) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 6 amino acids  
(B) TYPE: Amino Acid  
(D) TOPOLOGY: Linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:51:

His His His His His  
1 5 6

20 (2) INFORMATION FOR SEQ ID NO:52:

(i) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 4 amino acids  
(B) TYPE: Amino Acid  
(D) TOPOLOGY: Linear

25 (xi) SEQUENCE DESCRIPTION: SEQ ID NO:52:

Leu Met Arg Lys  
1 4

(2) INFORMATION FOR SEQ ID NO:53:

30 (i) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 4 amino acids  
(B) TYPE: Amino Acid  
(D) TOPOLOGY: Linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:53:

35 Leu Thr Ala Arg  
1 4

(2) INFORMATION FOR SEQ ID NO:54:

40 (i) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 4 amino acids  
(B) TYPE: Amino Acid  
(D) TOPOLOGY: Linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:54:

Gly Pro Gly Gly  
1 4

## (2) INFORMATION FOR SEQ ID NO:55:

5           (i) SEQUENCE CHARACTERISTICS:  
              (A) LENGTH: 5 amino acids  
              (B) TYPE: Amino Acid  
              (D) TOPOLOGY: Linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:55:

Gly Leu Met Arg Lys  
1                   5

## (2) INFORMATION FOR SEQ ID NO:56:

10           (i) SEQUENCE CHARACTERISTICS:  
              (A) LENGTH: 4 amino acids  
              (B) TYPE: Amino Acid  
              (D) TOPOLOGY: Linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:56:

15           Ala Ala Pro Phe  
1                   4

## (2) INFORMATION FOR SEQ ID NO:57:

20           (i) SEQUENCE CHARACTERISTICS:  
              (A) LENGTH: 13 amino acids  
              (B) TYPE: Amino Acid  
              (D) TOPOLOGY: Linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:57:

Gly Pro Gly Gly Xaa Xaa Xaa Xaa Xaa Gly Gly Pro Gly  
1                   5                   10                   13

## 25           (2) INFORMATION FOR SEQ ID NO:58:

(i) SEQUENCE CHARACTERISTICS:  
              (A) LENGTH: 4 amino acids  
              (B) TYPE: Amino Acid  
              (D) TOPOLOGY: Linear

30           (xi) SEQUENCE DESCRIPTION: SEQ ID NO:58:

Ala Ala Pro Lys  
1                   4

## (2) INFORMATION FOR SEQ ID NO:59:

35           (i) SEQUENCE CHARACTERISTICS:  
              (A) LENGTH: 4 amino acids  
              (B) TYPE: Amino Acid  
              (D) TOPOLOGY: Linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:59:

40           Ala Ala Pro Arg  
1                   4

## (2) INFORMATION FOR SEQ ID NO:60:

(i) SEQUENCE CHARACTERISTICS:  
              (A) LENGTH: 4 amino acids  
              (B) TYPE: Amino Acid

WO 96/27671

(D) TOPOLOGY: Linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:60:

Ala Ala Pro Met  
1 4

5 (2) INFORMATION FOR SEQ ID NO:61:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 4 amino acids
- (B) TYPE: Amino Acid
- (D) TOPOLOGY: Linear

10 (xi) SEQUENCE DESCRIPTION: SEQ ID NO:61:

Ala Ala Pro Gln  
1 4

(2) INFORMATION FOR SEQ ID NO:62:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 4 amino acids
- (B) TYPE: Amino Acid
- (D) TOPOLOGY: Linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:62:

20 Ala Ala Lys Phe  
1 4

(2) INFORMATION FOR SEQ ID NO:63:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 4 amino acids
- (B) TYPE: Amino Acid
- (D) TOPOLOGY: Linear

25 (xi) SEQUENCE DESCRIPTION: SEQ ID NO:63:

Ala Ala Ala Phe  
1 4

(2) INFORMATION FOR SEQ ID NO:64:

30 (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 4 amino acids
- (B) TYPE: Amino Acid
- (D) TOPOLOGY: Linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:64:

35 Ala Ala Arg Phe  
1 4

(2) INFORMATION FOR SEQ ID NO:65:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 4 amino acids
- (B) TYPE: Amino Acid
- (D) TOPOLOGY: Linear

40 (xi) SEQUENCE DESCRIPTION: SEQ ID NO:65:

Ala Ala Asp Phe

1 4

## (2) INFORMATION FOR SEQ ID NO:66:

5 (i) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 4 amino acids  
(B) TYPE: Amino Acid  
(D) TOPOLOGY: Linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:66:

Ala Ala Lys Lys  
1 4

## 10 (2) INFORMATION FOR SEQ ID NO:67:

(i) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 4 amino acids  
(B) TYPE: Amino Acid  
(D) TOPOLOGY: Linear

## 15 (xi) SEQUENCE DESCRIPTION: SEQ ID NO:67:

Ala Ala Lys Arg  
1 4

## (2) INFORMATION FOR SEQ ID NO:68:

20 (i) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 4 amino acids  
(B) TYPE: Amino Acid  
(D) TOPOLOGY: Linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:68:

Ala Ala Lys Phe  
25 1 4

## (2) INFORMATION FOR SEQ ID NO:69:

(i) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 4 amino acids  
(B) TYPE: Amino Acid  
30 (D) TOPOLOGY: Linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:69:

Ala Ala Pro Xaa  
1 4

## (2) INFORMATION FOR SEQ ID NO:70:

35 (i) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 4 amino acids  
(B) TYPE: Amino Acid  
(D) TOPOLOGY: Linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:70:

40 Ala Ala Xaa Phe  
1 4

## (2) INFORMATION FOR SEQ ID NO:71:

(i) SEQUENCE CHARACTERISTICS:

WO 96/27671

- (A) LENGTH: 5 amino acids
- (B) TYPE: Amino Acid
- (D) TOPOLOGY: Linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:71:

5	Ala Ala Xaa Xaa Xaa	
	1	5

## (2) INFORMATION FOR SEQ ID NO:72:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 275 amino acids
- (B) TYPE: Amino Acid
- (D) TOPOLOGY: Linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:72:

15	Ala Gln Ser Val Pro Tyr Gly Val Ser Gln Ile Lys Ala Pro Ala			
	1	5	10	15
15	Leu His Ser Gln Gly Tyr Thr Gly Ser Asn Val Lys Val Ala Val			
	20	25	30	
15	Ile Asp Ser Gly Ile Asp Ser Ser His Pro Asp Leu Lys Val Ala			
	35	40	45	
20	Gly Gly Ala Ser Met Val Pro Ser Glu Thr Asn Pro Phe Gln Asp			
	50	55	60	
20	Asn Asp Ser His Gly Thr His Val Ala Gly Thr Val Ala Ala Leu			
	65	70	75	
20	Asn Asn Ser Ile Gly Val Leu Gly Val Ala Pro Ser Ala Ser Leu			
	80	85	90	
25	Tyr Ala Val Lys Val Leu Gly Ala Asp Gly Ser Gly Gln Tyr Ser			
	95	100	105	
25	Trp Ile Ile Asn Gly Ile Glu Trp Ala Ile Ala Asn Asn Met Asp			
	110	115	120	
25	Val Ile Asn Met Ser Leu Gly Gly Pro Ser Gly Ser Ala Ala Leu			
	125	130	135	
25	Lys Ala Ala Val Asp Lys Ala Val Ala Ser Gly Val Val Val Val			
	140	145	150	
25	Ala Ala Ala Gly Asn Glu Gly Thr Ser Gly Ser Ser Ser Thr Val			
	155	160	165	
35	Asp Tyr Pro Gly Lys Tyr Pro Ser Val Ile Ala Val Gly Ala Val			
	170	175	180	
35	Asp Ser Ser Asn Gln Arg Ala Ser Phe Ser Ser Val Gly Pro Glu			
	185	190	195	
40	Leu Asp Val Met Ala Pro Gly Val Ser Ile Gln Ser Thr Leu Pro			
	200	205	210	
40	Gly Asn Lys Tyr Gly Ala Tyr Asn Gly Thr Ser Met Ala Ser Pro			
	215	220	225	
40	His Val Ala Gly Ala Ala Ala Leu Ile Leu Ser Lys His Pro Asn			
	230	235	240	

Trp Thr Asn Thr Gln Val Arg Ser Ser Leu Glu Asn Thr Thr Thr  
 245 250 255

Lys Leu Gly Asp Ser Phe Tyr Tyr Gly Lys Gly Leu Ile Asn Val  
 260 265 270

5 Gln Ala Ala Ala Gln  
 275

## (2) INFORMATION FOR SEQ ID NO:73:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 4 amino acids
- (B) TYPE: Amino Acid
- (C) STRANDEDNESS: Linear

10

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:73:

Arg Val Arg Arg  
 1 4

## 15 (2) INFORMATION FOR SEQ ID NO:74:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1146 base pairs
- (B) TYPE: Nucleic Acid
- (C) STRANDEDNESS: Single

20

## (D) TOPOLOGY: Linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:74:

G TG AGA GGC AAA AAA GTA TGG ATC AGT TTG CTG TTT 36  
 Val Arg Gly Lys Lys Val Trp Ile Ser Leu Leu Phe  
 -107 -105 -100

25

GCT TTA GCG TTA ATC TTT ACG ATG GCG TTC GGC AGC ACA 75  
 Ala Leu Ala Leu Ile Phe Thr Met Ala Phe Gly Ser Thr  
 -95 -90 -85

30

TCC TCT GCC CAG GCG GCA GGG AAA TCA AAC GGG GAA AAG 114  
 Ser Ser Ala Gln Ala Ala Gly Lys Ser Asn Gly Glu Lys  
 -80 -75 -70

AAA TAT ATT GTC GGG TTT AAA CAG ACA ATG AGC ACG ATG 153  
 Lys Tyr Ile Val Gly Phe Lys Gln Thr Met Ser Thr Met  
 -65 -60

35

AGC GCC GCT AAG AAG AAA GAT GTC ATT TCT GAA AAA GGC 192  
 Ser Ala Ala Lys Lys Asp Val Ile Ser Glu Lys Gly  
 -55 -50 -45

GGG AAA GTG CAA AAG CAA TTC AAA TAT GTA GAC GCA GCT 231  
 Gly Lys Val Gln Lys Gln Phe Lys Tyr Val Asp Ala Ala  
 -40 -35

40

TCA GCT ACA TTA AAC GAA AAA GCT GTA AAA GAA TTG AAA 270  
 Ser Ala Thr Leu Asn Glu Lys Ala Val Lys Glu Leu Lys  
 -30 -25 -20

45

AAA GAC CCG AGC GTC GCT TAC GTT GAA GAA GAT CAC GTA 309  
 Lys Asp Pro Ser Val Ala Tyr Val Glu Glu Asp His Val  
 -15 -10 -5

AGA CAT AAG CGC GCG CAG TCC GTG CCT TAC GGC GTA TCA 348

WO 96/27671

Arg His Lys Arg Ala Gln Ser Val Pro Tyr Gly Val Ser  
1 5

CAA ATT AAA GCC CCT GCT CTG CAC TCT CAA GGC TAC ACT 387  
Gln Ile Lys Ala Pro Ala Leu His Ser Gln Gly Tyr Thr  
5 10 15 20

GGA TCA AAT GTT AAA GTA GCG GTT ATC GAC AGC GGT ATC 426  
Gly Ser Asn Val Lys Val Ala Val Ile Asp Ser Gly Ile  
25 30 35

GAT TCT TCT CAT CCT GAT TTA AAG GTA GCA GGC GGA GCC 465  
Asp Ser Ser His Pro Asp Leu Lys Val Ala Gly Gly Ala  
10 40 45

AGC ATG GTT CCT TCT GAA ACA AAT CCT TTC CAA GAC AAC 504  
Ser Met Val Pro Ser Glu Thr Asn Pro Phe Gln Asp Asn  
50 55 60

GAC TCT CAC GGA ACT CAC GTT GCC GGC ACA GTT GCG GCT 543  
Asp Ser His Gly Thr His Val Ala Gly Thr Val Ala Ala  
15 65 70

CTT AAT AAC TCA ATC GGT GTA TTA GGC GTT GCG CCA AGC 582  
Leu Asn Asn Ser Ile Gly Val Leu Gly Val Ala Pro Ser  
20 75 80 85

GCA TCA CTT TAC GCT GTA AAA GTT CTC GGT GCT GAC GGT 621  
Ala Ser Leu Tyr Ala Val Lys Val Leu Gly Ala Asp Gly  
90 95 100

TCC GGC CAA GAT AGC TGG ATC ATT AAC GGA ATC GAG TGG 660  
Ser Gly Gln Asp Ser Trp Ile Ile Asn Gly Ile Glu Trp  
25 105 110

GCG ATC GCA AAC AAT ATG GAC GTT ATT AAC ATG AGC CTC 699  
Ala Ile Ala Asn Asn Met Asp Val Ile Asn Met Ser Leu  
115 120 125

GGC GGA CCT TCT GGT TCT GCT GCT TTA AAA GCG GCA GTT 738  
Gly Gly Pro Ser Gly Ser Ala Ala Leu Lys Ala Ala Val  
30 130 135

GAT AAA GCC GTT GCA TCC GGC GTC GTA GTC GTT GCG GCA 777  
Asp Lys Ala Val Ala Ser Gly Val Val Val Val Ala Ala  
35 140 145 150

GCC GGT AAC GAA GGC ACT TCC GGC AGC TCG TCG ACA GTG 816  
Ala Gly Asn Glu Gly Thr Ser Gly Ser Ser Thr Val  
155 160 165

GAC TAC CCT GGC AAA TAC CCT TCT GTC ATT GCA GTA GGC 855  
Asp Tyr Pro Gly Lys Tyr Pro Ser Val Ile Ala Val Gly  
40 170 175

GCT GTT GAC AGC AGC AAC CAA AGA GCA TCT TTC TCA AGC 894  
Ala Val Asp Ser Ser Asn Gln Arg Ala Ser Phe Ser Ser  
180 185 190

GTA GGA CCT GAG CTT GAT GTC ATG GCA CCT GGC GTA TCT 933  
Val Gly Pro Glu Leu Asp Val Met Ala Pro Gly Val Ser  
45 195 200

ATC CAA AGC ACG CTT CCT GGA AAC AAA TAC GGG GCG TAC 972  
Ile Gln Ser Thr Leu Pro Gly Asn Lys Tyr Gly Ala Tyr

205	210	215
AAC GGT ACC TCA ATG GCA TCT CCG CAC GTT GCC GGA GCG 1011 Asn Gly Thr Ser Met Ala Ser Pro His Val Ala Gly Ala 220 225 230		
5	GCT GCT TTG ATT CTT TCT AAG CAC CCG AAC TGG ACA AAC 1050 Ala Ala Leu Ile Leu Ser Lys His Pro Asn Trp Thr Asn 235 240	
10	ACT CAA GTC CGC AGC AGT TTA GAA AAC ACC ACT ACA AAA 1089 Thr Gln Val Arg Ser Ser Leu Glu Asn Thr Thr Lys 245 250 255	
CTT GGT GAT TCT TTC TAC TAT GGA AAA GGG CTG ATC AAC 1128 Leu Gly Asp Ser Phe Tyr Tyr Gly Lys Gly Leu Ile Asn 260 265		
15	GTA CAG GCG GCA GCT CAG 1146 Val Gln Ala Ala Ala Gln 270 275	
(2) INFORMATION FOR SEQ ID NO:75:		
(i) SEQUENCE CHARACTERISTICS:		
20	(A) LENGTH: 382 amino acids (B) TYPE: Amino Acid (C) TOPOLOGY: Linear	
(xi) SEQUENCE DESCRIPTION: SEQ ID NO:75:		
Val Arg Gly Lys Lys Val Trp Ile Ser Leu Leu Phe Ala Leu Ala -107 -105 -100 -95		
25	Leu Ile Phe Thr Met Ala Phe Gly Ser Thr Ser Ser Ala Gln Ala -90 -85 -80	
Ala Gly Lys Ser Asn Gly Glu Lys Lys Tyr Ile Val Gly Phe Lys -75 -70 -65		
30	Gln Thr Met Ser Thr Met Ser Ala Ala Lys Lys Lys Asp Val Ile -60 -55 -50	
Ser Glu Lys Gly Lys Val Gln Lys Gln Phe Lys Tyr Val Asp -45 -40 -35		
Ala Ala Ser Ala Thr Leu Asn Glu Lys Ala Val Lys Glu Leu Lys -30 -25 -20		
35	Lys Asp Pro Ser Val Ala Tyr Val Glu Glu Asp His Val Arg His -15 -10 -5	
Lys Arg Ala Gln Ser Val Pro Tyr Gly Val Ser Gln Ile Lys Ala 1 5 10		
40	Pro Ala Leu His Ser Gln Gly Tyr Thr Gly Ser Asn Val Lys Val 15 20 25	
Ala Val Ile Asp Ser Gly Ile Asp Ser Ser His Pro Asp Leu Lys 30 35 40		
Val Ala Gly Gly Ala Ser Met Val Pro Ser Glu Thr Asn Pro Phe 45 50 55		
45	Gln Asp Asn Asp Ser His Gly Thr His Val Ala Gly Thr Val Ala	

WO 96/27671

	60	65	70
	Ala Leu Asn Asn Ser Ile Gly Val Leu Gly Val Ala Pro Ser Ala		
	75	80	85
5	Ser Leu Tyr Ala Val Lys Val Leu Gly Ala Asp Gly Ser Gly Gln		
	90	95	100
	Asp Ser Trp Ile Ile Asn Gly Ile Glu Trp Ala Ile Ala Asn Asn		
	105	110	115
	Met Asp Val Ile Asn Met Ser Leu Gly Gly Pro Ser Gly Ser Ala		
	120	125	130
10	Ala Leu Lys Ala Ala Val Asp Lys Ala Val Ala Ser Gly Val Val		
	135	140	145
	Val Val Ala Ala Ala Gly Asn Glu Gly Thr Ser Gly Ser Ser Ser		
	150	155	160
15	Thr Val Asp Tyr Pro Gly Lys Tyr Pro Ser Val Ile Ala Val Gly		
	165	170	175
	Ala Val Asp Ser Ser Asn Gln Arg Ala Ser Phe Ser Ser Val Gly		
	180	185	190
	Pro Glu Leu Asp Val Met Ala Pro Gly Val Ser Ile Gln Ser Thr		
	195	200	205
20	Leu Pro Gly Asn Lys Tyr Gly Ala Tyr Asn Gly Thr Ser Met Ala		
	210	215	220
	Ser Pro His Val Ala Gly Ala Ala Ala Leu Ile Leu Ser Lys His		
	225	230	235
25	Pro Asn Trp Thr Asn Thr Gln Val Arg Ser Ser Leu Glu Asn Thr		
	240	245	250
	Thr Thr Lys Leu Gly Asp Ser Phe Tyr Tyr Gly Lys Gly Leu Ile		
	255	260	265
	Asn Val Gln Ala Ala Ala Gln		
	270	275	

## 30 (2) INFORMATION FOR SEQ ID NO:76:

- (i) SEQUENCE CHARACTERISTICS:
  - (A) LENGTH: 5 amino acids
  - (B) TYPE: Amino Acid
  - (D) TOPOLOGY: Linear

## 35 (xi) SEQUENCE DESCRIPTION: SEQ ID NO:76:

Asn Arg Met Arg Lys	
1	5

## (2) INFORMATION FOR SEQ ID NO:77:

- (i) SEQUENCE CHARACTERISTICS:
  - (A) LENGTH: 11 amino acids
  - (B) TYPE: Amino Acid
  - (D) TOPOLOGY: Linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:77:

Gly Ser Gly Gln Tyr Ser Trp Ile Ile Asn Gly  
1 5 10 11

## (2) INFORMATION FOR SEQ ID NO:78:

5 (i) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 11 amino acids  
(B) TYPE: Amino Acid  
(D) TOPOLOGY: Linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:78:

Gly Asp Ile Thr Thr Glu Asp Glu Ala Ala Ser  
10 1 5 10 11

## (2) INFORMATION FOR SEQ ID NO:79:

15 (i) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 11 amino acids  
(B) TYPE: Amino Acid  
(D) TOPOLOGY: Linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:79:

Gly Glu Val Thr Asp Ala Val Glu Ala Arg Ser  
1 5 10 11

## (2) INFORMATION FOR SEQ ID NO:80:

20 (i) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 11 amino acids  
(B) TYPE: Amino Acid  
(D) TOPOLOGY: Linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:80:

25 Pro Phe Met Thr Asp Ile Ile Glu Ala Ser Ser  
1 5 10 11

## (2) INFORMATION FOR SEQ ID NO:81:

30 (i) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 11 amino acids  
(B) TYPE: Amino Acid  
(D) TOPOLOGY: Linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:81:

Gly Ile Val Thr Asp Ala Ile Glu Ala Ser Ser  
1 5 10 11

## 35 (2) INFORMATION FOR SEQ ID NO:82:

(i) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 27 base pairs  
(B) TYPE: Nucleic Acid  
(C) STRANDEDNESS: Single  
(D) TOPOLOGY: Linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:82:

GGTTCCGGCC AAGATAGCTG GATCATT 27

## (2) INFORMATION FOR SEQ ID NO:83:

5 (i) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 29 base pairs  
(B) TYPE: Nucleic Acid  
(C) STRANDEDNESS: Single  
(D) TOPOLOGY: Linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:83:

CCAATACAGC TGGAAATTA ACGGAATCG 29

(2) INFORMATION FOR SEQ ID NO:84:

10 (i) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 31 base pairs  
(B) TYPE: Nucleic Acid  
(C) STRANDEDNESS: Single  
(D) TOPOLOGY: Linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:84:

15 GGTTCGGGCC AAGATAGCTG GGAAATTAAAC G 31

(2) INFORMATION FOR SEQ ID NO:85:

20 (i) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 30 base pairs  
(B) TYPE: Nucleic Acid  
(C) STRANDEDNESS: Single  
(D) TOPOLOGY: Linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:85:

AAGAAGATCA CGTAAGACAT AAGCGCGCGC 30

(2) INFORMATION FOR SEQ ID NO:86:

25 (i) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 4 amino acids  
(B) TYPE: Amino Acid  
(D) TOPOLOGY: Linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:86:

30 Arg Ala Lys Arg  
1 4

(2) INFORMATION FOR SEQ ID NO:87:

35 (i) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 4 amino acids  
(B) TYPE: Amino Acid  
(D) TOPOLOGY: Linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:87:

Lys Ala Lys Arg  
1 4

40 (2) INFORMATION FOR SEQ ID NO:88:

(i) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 8 amino acids

WO 96/27671

PCT/US96/02861

(B) TYPE: Amino Acid  
(D) TOPOLOGY: Linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:88:

5 Gly Pro Gly Gly Leu Met Arg Lys  
1 5 8

(2) INFORMATION FOR SEQ ID NO:89:

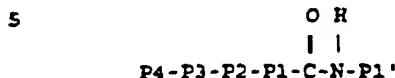
(i) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 8 amino acids  
(B) TYPE: Amino Acid  
10 (D) TOPOLOGY: Linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:89:

Gly Pro Gly Gly Lys Ala Lys Arg  
1 5 8

What is claimed is:

1. A subtilisin variant derived from a precursor subtilisin-type serine protease said variant capable of cleaving a polypeptide substrate comprising the sequence:



wherein:

P4 is a basic amino acid;

10 P3 is any amino acid selected from the naturally occurring amino acids;

P2 is a basic amino acid;

P1 is a basic amino acid; and

P1' is not Pro.

2. The subtilisin variant of claim 1 containing an acidic amino acid at a residue equivalent to

15 Asn 62, Tyr 104 and Gly 166 of the subtilisin naturally produced by *Bacillus amyloliquefaciens*.

3. The subtilisin-type serine protease variant of claim 2 wherein the acidic amino acid is Asp or Glu.

4. The subtilisin-type serine protease variant of claim 3 wherein the acidic amino acid is Asp.

5. The subtilisin-type serine protease variant of claim 2 wherein the precursor subtilisin-type 20 serine protease in the subtilisin naturally produced by *Bacillus amyloliquefaciens*.

6. The subtilisin variant of claim 5 having the amino acid sequence of the mature polypeptide of Figure 8 (SEQ ID NO: 75).

7. A subtilisin variant having substrate specificity for peptide substrates containing dibasic amino acid sequences.

25 8. The subtilisin variant of claim 7 having a different amino acid residue at residue position +62 than subtilisin naturally produced by *Bacillus amyloliquefaciens*.

9. The subtilisin variant of Claim 8 having an Asp or Glu at residue position +62.

10. The subtilisin variant of Claim 9 having an Asp at residue position +62.

11. The subtilisin variant of Claim 10 further having an Asp or Glu at residue position +166.

30 12. The subtilisin variant of Claim 11 having an Asp at residue position +166.

13. The subtilisin variant of Claim 12 having the amino acid sequence of the mature polypeptide provided in Fig. 6.

14. An isolated nucleic acid molecule encoding the subtilisin variant of Claim 1.

15. The nucleic acid molecule of Claim 14 further comprising a promoter operably linked to the 35 nucleic acid molecule.

16. An expression vector comprising the nucleic acid molecule of Claim 15 operably linked to control sequences recognized by a host cell transformed with the vector.

17. A host cell transformed with the vector of Claim 16.

18. An isolated nucleic acid molecule encoding the subtilisin variant of Claim 7.

19. The nucleic acid molecule of Claim 18 further comprising a promoter operably linked to the nucleic acid molecule.

20. An expression vector comprising the nucleic acid molecule of Claim 19 operably linked to control sequences recognized by a host cell transformed with the vector.

5 21. A host cell transformed with the vector of Claim 20.

22. A process of using the nucleic acid molecule encoding the subtilisin variant to effect production of the subtilisin variant comprising culturing the host cell of Claim 21 under conditions suitable for expression of the subtilisin variant.

10 23. The process of Claim 22 further comprising recovering the subtilisin variant from the host cell culture medium.

24. A method of using the subtilisin variant of Claim 1 comprising contacting a fusion protein containing a dibasic sequence with the subtilisin variant.

25. A process for cleaving a polypeptide, said polypeptide comprising an amino acid sequence represented by the formula:

15 P4-P3-P2-P1-P1'

wherein,

P4 is a basic amino acid;

P3 is an amino acid selected from the naturally occurring amino acids;

P2 is a basic amino acid;

20 P1 is a basic amino acid; and

P1' is not Pro;

comprising the step of:

subjecting said polypeptide to the subtilisin variant of claim 1 in a reaction mixture under conditions such that the subtilisin variant cleaves the polypeptide.

25 26. A process of using the nucleic acid molecule encoding the subtilisin variant to effect production of the subtilisin variant comprising culturing the host cell of Claim 17 under conditions suitable for expression of the subtilisin variant.

27. The process of Claim 26 further comprising recovering the subtilisin variant from the host cell culture medium.

30 28. A method of using the subtilisin variant of Claim 7 comprising contacting a fusion protein containing a dibasic sequence with the subtilisin variant.

29. A process for cleaving a polypeptide, said polypeptide comprising an amino acid sequence represented by the formula:

P4-P3-P2-P1-P1'

35 wherein,

P4 is a large hydrophobic amino acid;

P3 is an amino acid selected from the naturally occurring amino acids;

P2 is a basic amino acid;

P1 is a basic amino acid; and

WO 96/27671

PCT/US96/02861

P1' is not Pro;  
comprising the step of:  
subjecting said polypeptide to the subtilisin variant of claim 7 in a reaction mixture under conditions  
such that the subtilisin variant cleaves the polypeptide.

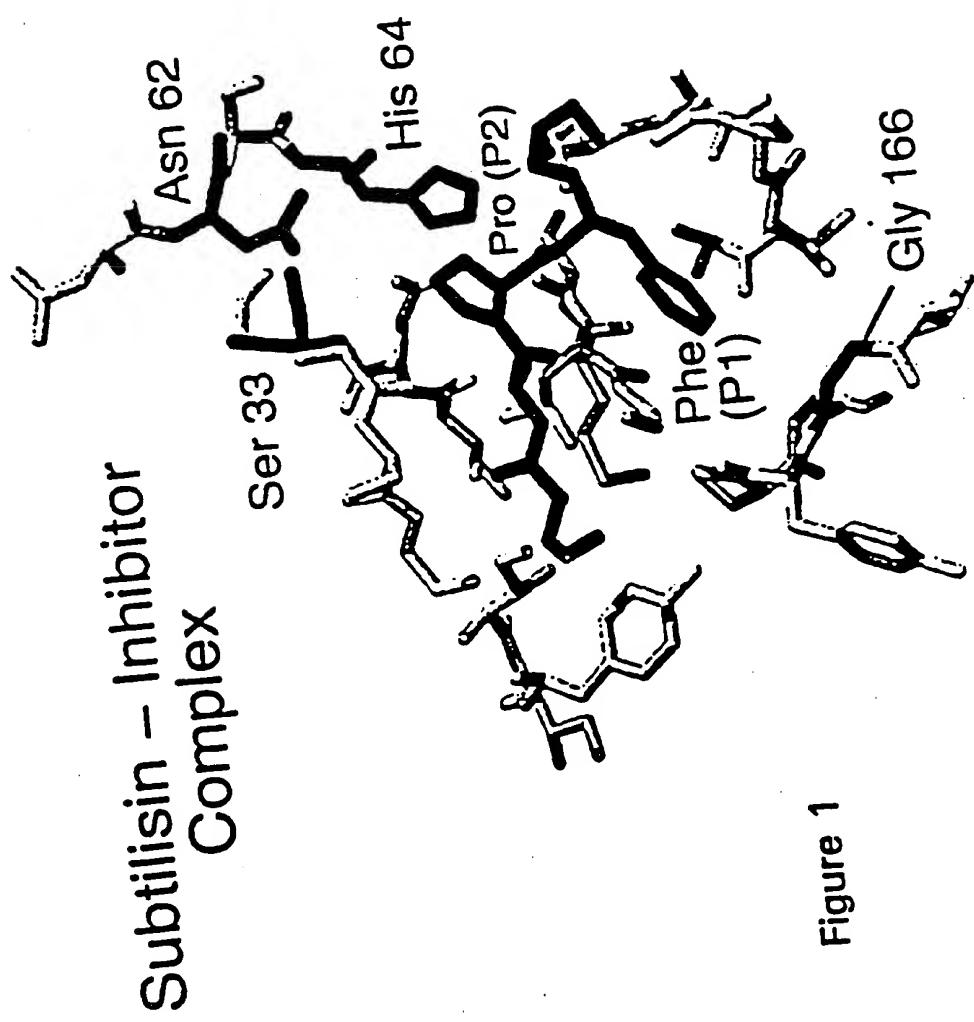


Figure 1

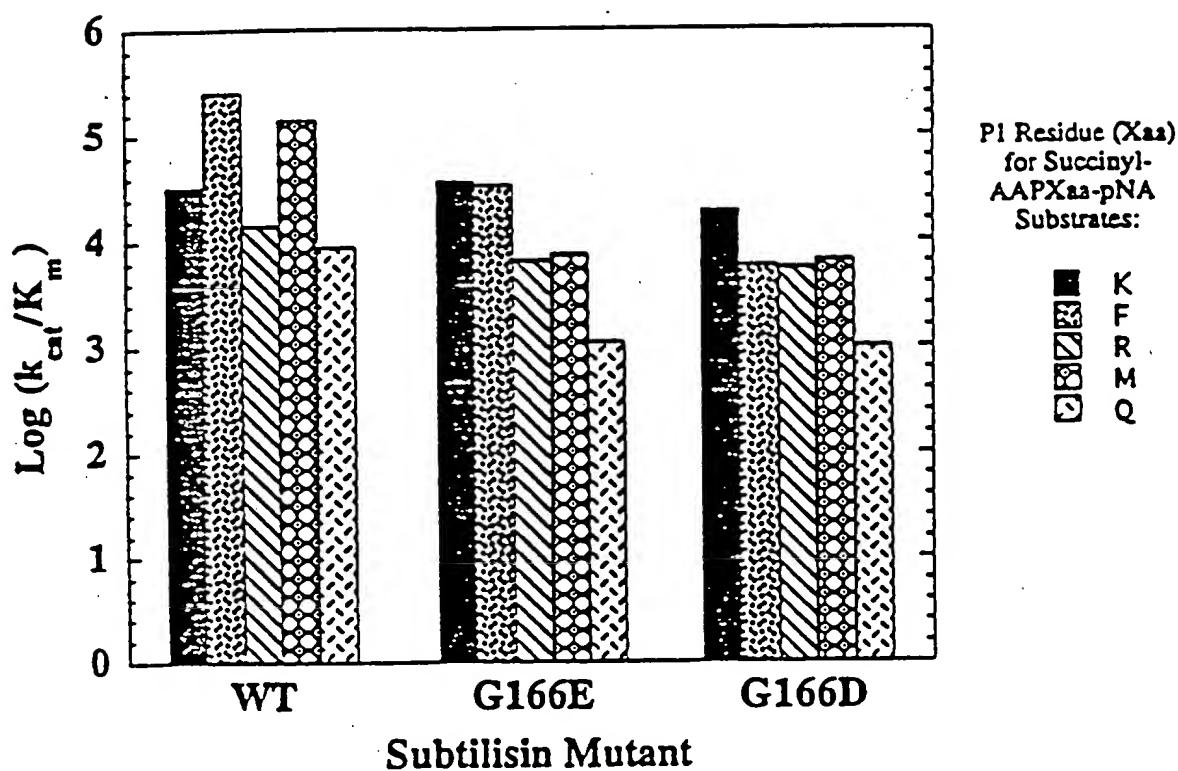


Figure 2

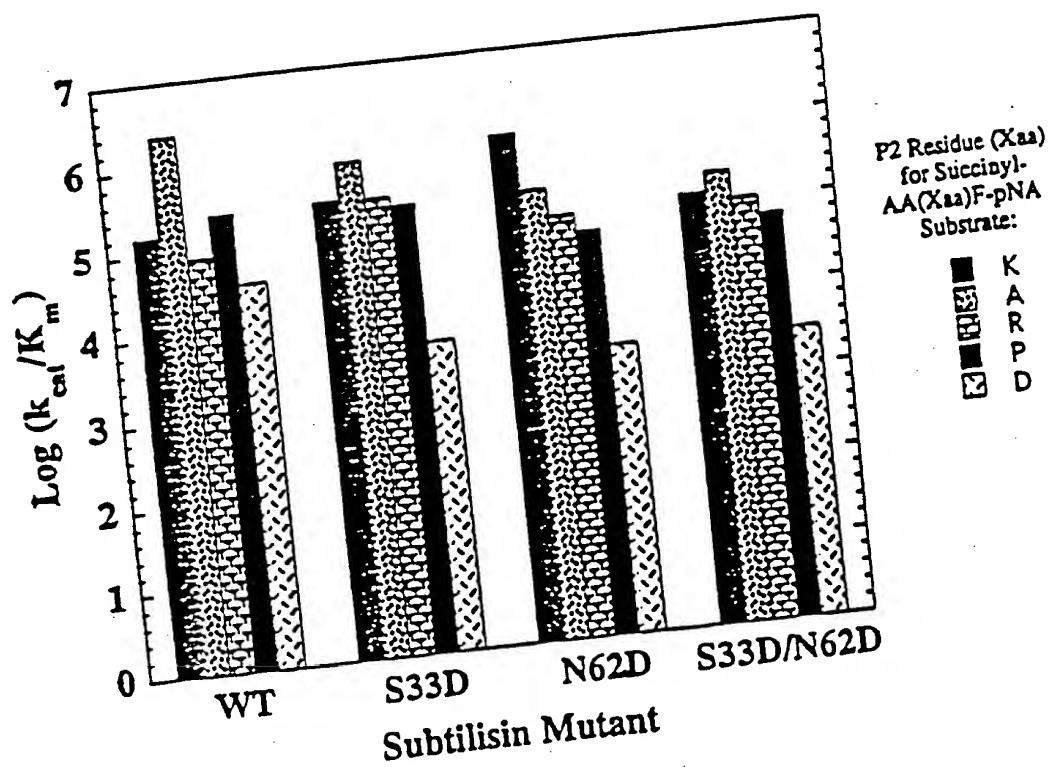


Figure 3

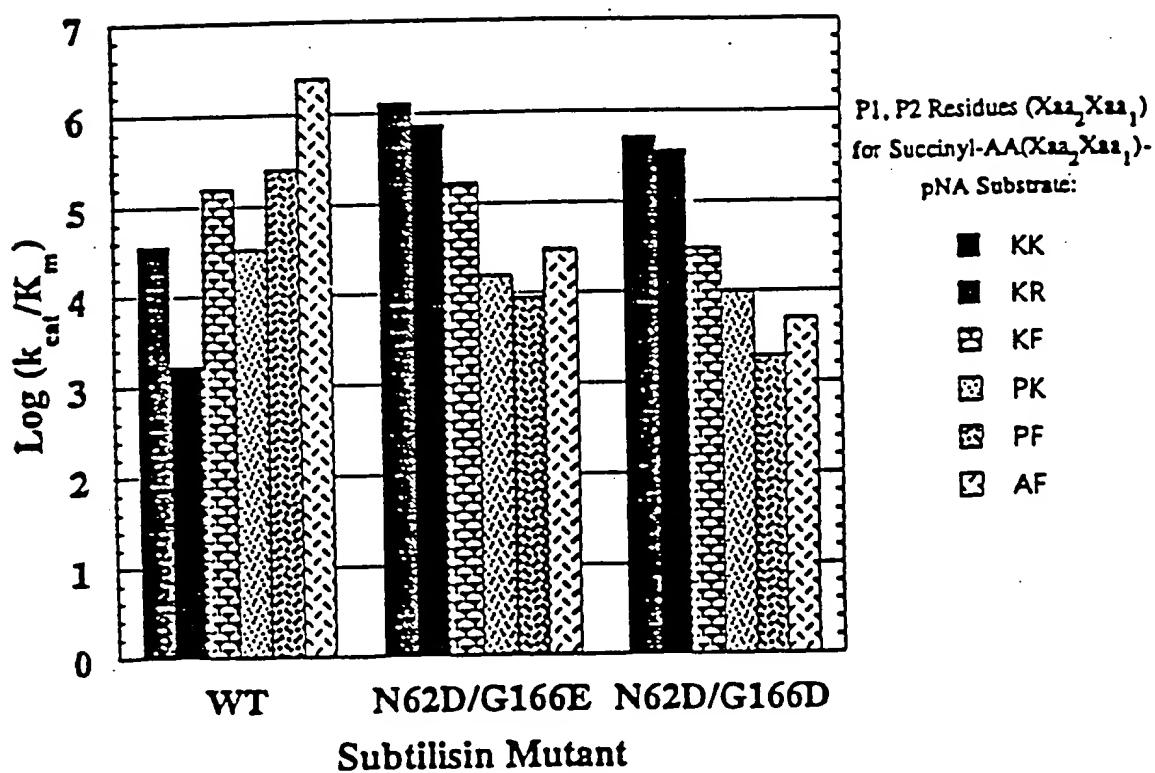


Figure 4

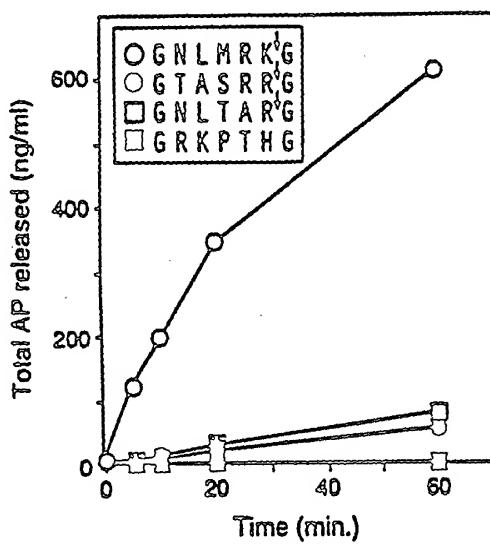


FIGURE 5

**Figure 6-1**

PCT/US96/02861

7/17

WO 96/27671

Figure 6-2

WO 96/27671

8/17

**Figure 6-3**

9/17

Figure 6-4

WO 96/27671

10/17

ecos71

**2921** AGAGGGAGCT ATGAGGGGG TGTACAGAG TCTTGAGT GGTGGCTTA CTACCGTAC AGCAGAGGA CAGATTGG TGTCGGCT CGCTGAGC  
TCTCGCTCCA TACATCCGGC ACCATGTCTC AAGAACTCA CCACGGATT GATGCCATG TGCTTCTT GTCATMAGC ATAGACGGG GACGACTTC

**3021** CAGTTACCTT CGGAAAAGA GTGGTAGCT CTGATCCGG CAACAAAC ACCGGCTGA CGCGCTGTT TTGCTTCG AAGGGAGA TAGGGAG  
TCTCGCTCCA GCCTTTCT CAACATCGA GAACTAGGC GTTGTGTT GGCGACCAT CGCCACCAA AAACAAAG TTGGCTGCT ATAGGGGTC

Fpol

**3121** AAAAGGAGCT TCACCTAGAT CCTTTAAAT TAAATGAA GTTAAATC AACCTAACCT ATATAGACT AACCTGGC TGACAGTC CAACTCTTA  
TTTCTTCAAG AGCTACCTA GAAATTTA ATTTCAGT CAAATTCG TTGATTCG TATATACCA TTATACCA AACCTGTA CTCATAGT

BpuI

**3221** TCACTGAGGC ACCTATCTCA CGGATCTGTC TATTCGTC ATCCAGCT CCTGACTCC CGGATCTCA GAACTGAGG ACTGCTCTG GTTACGATT  
AGTCACTCCG TGGATAGCT CGCTAGACAG ATAAAGCAG TGGATTCG CGGAGCACAT CTATGAGC GGTACCTC CGATGGTAG

HincII

**3321** TCCCCCAGT CCTGCATCA TACCGAGA CCCACGCTCA CGGGCTCCAG ATTATCAGC AACAAACAG CCTGGCTCC TATGGCTC TATGGCTC  
ACCGGGTCA CGACCTACT ATGGGCTCT GGGTGGAGT GGGCGGGTC TAAATGAGC TTATGGCT CCTGGCTT CCTGGCTCC GTCCTCACCA

SmaI/SphI

**3421** CCTGGCAGCTT TATCCGGCTC CATCCAGCTT ATTATGCTT CGGGGAGC TAGCTGAGT AGTGGCCAG TATATGCTT GGGCGGGTC  
GGACCGTCA ATAGGGGAGC GTAGGTAGA TAATACCA CGGCCCTCG ATCCATCA TCAAGGGGTC AACATCAA CGCGTCCAA CAGGGTAC

BglI

**3521** CGCGGAGAT CCGATGCA GCGATGCTT TGGATGGG TGTATGAG TGGATGGC AACCTACAG GCGAGCTCA TGTATGAGA TGTATGAG  
GACGCGCTA GCACGAGT CGCAGGAGCA ACCATACCG AGCTAGCTG AGGCCAGGG TTGCTGAGTC CGCTCAATGT ACTAGGGT AACACAGCTT

PvuI/BpuI

**3621** AAAAGGCGTT AGCTTCTCA CGCTCGAT CCTTGTGAG ACTAATGGA CGCGAGCTT ATCCATCA GGTAGGGAG CACTGATA TCTCTACT  
TTTCGCCAA TCGAGGAGC CAGGAGGCTA GCAACAGCTT TCAATCACCA TGGTGTAGTC CAACTGGTC GTGACGGATT AACGAGATCA

BpuI

BpuI

BpuI

Figure 6-5

WO 96127671

11 / 17

**Figure 6-6**

WO 96/27671

12/17

XbaI  
Asp700

1421 AACGGAGG TAAAGAAC ACCAATCGA GATCAGGAA TGACTTATA AAATAAAAGGACACTGA AGGTGCTT TTTCGATG TTTGACTT  
TGGCCTCTCC ATTTTTCT TGCTCAGCT CTAGCCCT ACTCAATAA TTATTTT CGGGACTT TTCCACAGA AAACATAC AACATGAA

1521 GTTCTTCTT ATCTGATAC ATATAGAAAT AACGTCATT TTATTTGT TGCTGAAGG TCGGGTAG TATGCTTT AACGATCA  
CAAGAAAGA TAGAATAG TATCTTA TTGGCTAA ATAAATCA AGCACTTC AGCAACTTC AGCACATAC ATACACAAA TTCAATAC

1621 AAACCTTAA ATGGGTGC ACAGAMAC CCCATCGTT AAATTTAA GTGACTAAC AAATCTAA ATAGATGGG CTTCTTA ATATATGTC  
TTGGAAATT TAACCAGG TGTCTTTG GGCTAGACA TTCAATATT CACTGATTG TTATGATT TTCTACCC CAAGAAAT TATAAAC

BamHII/BsrI  
SstI  
AccI

1721 TCCTTAACT AGCATTATT CAGATGAAA ATCAAGGTT TTACTGGACA AGACAAAAG TCGAAAAGTG AGACCATGGA GAGAAAGAA ATCGCTTAA  
AGGATTATCA TCGTAATTA GTCTACTTT TTGTTCCAA ATTCACCTGT TCTTTTTC AGCTTCAC TCTGCTACCT CTCCTTC ATTGCAAT

BsrI/BsrI  
SspI  
AccI

1821 CTGCTTAACT TTGACCTCT GCATTTCTT GATTTAA AGCTGAAG AGTAAAGAT TTGCTGAA TATAGAGTA TAAACAAAT CGAGAACAG  
CACTAATCA AACTGAGA CGTATAGAA CTAAATTT TCGGACTTTC TCAATTCTA AGCAACTTT ATATCTCA ATTGTCTTA GCACTTTC

XbaI  
SspI  
AccI

1921 CGCAAGAGAA GTGTATCGA GTGGGTTT GAAATCGA CCTTGCTCA ATGCGAACT GGAGAGGG ATGCGAACT GGAGAGGT  
CGCTTCTT CAACTAGCT CACACCAAAT CTTAGGTC CGAACAGGT TACCGTTG CCTCTCTG TTACTTGTG CGTAGTCG GCTTTCCA

Eco571  
XbaI

2021 TGTGGCTGAA GTTAAAC AAAAGCCAC ACTTGTTG TCTTTCTCA CATTAGT TAAATGT TATGATGGG AGATTAAGA TAAAGTTG  
ACACGACTT CAAATTC TTGGTGTG TCAACCAAC AACAGAGT TAAATCTCA ATTTCACAA ATACTACCG TTCTTATT ATTCCAAAC

BsrI/ApaI/YspI  
AccI/ApaI/YspI

2121 TCAGATATG CTCAGGATT TGGCGAATG ATCCATATA AAAATTA TAAATCTT GTGGTTTA TGGCTCCAC CGACTGACA ATAAATATA  
AGCTATACG GAGTCTCAA AGGGCTTAC TACCTTAT TTCTTATT ATTTAGAA CAACCAAAAT AGCACGTT CCTTACTGT TATTATAT

BsrI  
ApaI  
Ppu10I  
BsrI/Avall

BsrI  
ApaI  
Ppu10I  
BsrI/Avall

2221 AGATAATTC TATATCGA CACTGATG TATGGGATG TGTGGACCA ATTTTTA AGATACAGA AAATCTGCA ATGAAACAC ATGGATCA  
TTCTTAACT AGTATAGTC GTGAGCTAC ATAACATAC ACACCTGGT TGAATTAAT TCTTATGCT TTGATGGAC TTCTTGT TTACCTAAT

Figure 6-7

Figure 6-8

14/17

6221 AGCTATCAA ACTGCTTAAAT CGGTAGAAC CCATAACCTTC CACCGATGCA TTTCGGCCT TATGGAGA GACGGCTTC ATATCCAGG CAATGGCT  
 TCACATAGT TGACGATTG CCATCTTCG GGTTCCAG GTCCTACGTT AACACGGCA ATAGCATCTT CTGCACAAAC TTATGGTC GTTAACCGCA  
**sp1**  
 6321 AAATTCGTC TGCAGGGACC GACACATT CTACCATCTT TGACTGCTC bsp1407I bsp1407I hglCI  
 TTATACAC ACGGTCCCTGG CTGTGTAAG ATGGCTAGGA ACTGACATG CCATGTTAC CGTCCACGGT ACACTAACC AGACGTGA CGGAGACAA  
**hsaII**  
 ec007III ec007III bsrII/xbaII  
 ec007III ec007III bglII ec007III bsp1407I bsp1407I baiI  
 6421 ATTCGAGGAG CGCTCGGTC TAACTGAG CACTAACCA ATCAGATCTT CTCAGGTT ATGACCATCT GTCCAGCTC GTCAGCTC GTCAGCTC  
 TATGGCTC GCGAAGCCAG ATAGACTTC GTCAGCTC GTCAATCGT TACTCTGAA GGAGCTGCA TACTGGTCA CACGCCAG CATTACGAC CAGTGAGA  
**pol1**  
 6521 CGACTCTGAG AACTCTGG ATTCGCTGA GAATTCCTGG ATGGGATC AGGAGGAC AGAACACAG GGTATATAG TGGATGTC AAAACGATA  
 GCTGAGACTC TTGAGAGCC TTAGGATCT CTAAAGCC TTACCTTAG TCTTCACCTG TCTTCCTG CTCATATAC ACCTACACG CAGTGAGA  
**bsaBI**  
 6621 CCATTTGAA CGATGACCTC TAAATGGT TAACTGTT GCTTCGGTAT TATTACCT CTCCAGTAT TACCTATC CTCGGCTCTC ATGGGCATT  
 GCTAAACTT GCTACTGGAG ATTAAACA ATTACCAA CCAATGATA ATTAAAGAA GAGGATATA ATCAATAA GTACCGACAG TTTGGGTAA  
**bsaI**  
 6721 AACGGATAA AGGGTGTCT TAATGGCC CATTGGGT ATAGAGAA AGGATTTT ATGGGATC AGGAGGAT TAAAGGTTAT AGATTACAT  
 TTGGCTTATT TCCCACACCA ATTAGCCCG CTAAACCA TATCTTCTT TCCATTTA TACCTGTTA ACTTAACTT TACCCATT TCAATGTA  
**bsbI/bsp1/vsp1**  
 6821 TAGAAATGA AGGGGATT TATGGTGAG ATGTTACAG TCAATCCGG CTAATCTAC AGCTAGGAA AGAAAGGAT TTTGGCTA  
 ATCTTTACT TTCCCTTAAT ATACGACTC TTACATCTC AGTATGGCC GTTATCTAG GGATATAG TTCTATCTT TTCTTCTA AAAGGCTG  
**shalII/draI**  
 6921 GCTCAATCC TTTAAAGAA CACAAAGAC CACATTTT ATCTGGCT TATTTCTCA ACTAAACAC CCATTTCTC AAGAACGAA ATGGATA  
 CGAGCTTGG AAATTTTT GTGTTCTG GTGAAAGAA TTACACCGA ATATAGCT TATTTCTG GGTATCGAG TTCTTCTT AAAGGCTG  
**sp1**  
 7021 ACTGGGATAT TTTAAAGAA TATTTCTG TTACGATAT ATTCAGCTT AAAGAGGAT TGATCTTAAT GAGAAAGCA GACAGTAAAG CCTCTCTAAAT  
 ATGAAATCT ATTTTAAAT CCTCCGATA CTTCATGAA ATATTTA ACTATCTG TTACCTCTT CTTCCTCA TAAATGTA ATATCTG  
**pol**  
 7121 TCACTTAAAGA TAAATTTA GAGGGATAT CAAATGACT TAAATTAAT TGATTTAGC AATTCGAGA GAAAGAT ATTATCAT TATTCGAA  
 ATGAAATCT ATTTTAAAT CCTCCGATA CTTCATGAA ATATTTA ACTATCTG TTACCTCTT CTTCCTCA TAAATGTA ATATCTG  
**bunI**  
 7221 ACTGGGATAT TTTAAAGAA TATTTCTG TTACGATAT ATTCAGCTT AAAGAGGAT TGATCTTAAT GAGAAAGCA GACAGTAAAG CCTCTCTAAAT  
 ATGAAATCT ATTTTAAAT CCTCCGATA CTTCATGAA ATATTTA ACTATCTG TTACCTCTT CTTCCTCA TAAATGTA ATATCTG

**Figure 6-9**

WD 96/27671

15/17

**Figure 6-10**

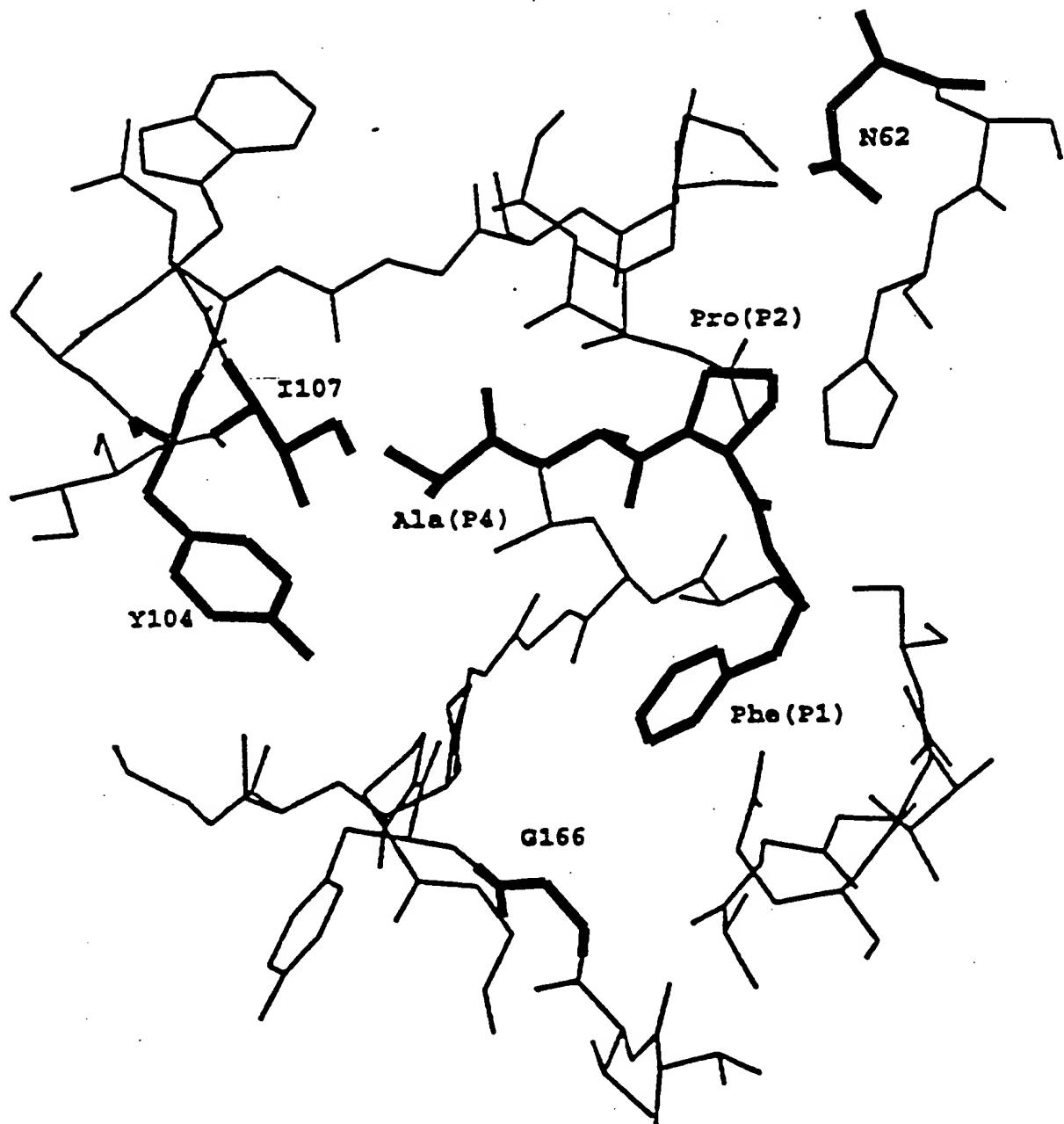


FIGURE 7

FIGURE 8

Val Arg Gly Lys Lys Val Trp Ile Ser Leu Leu Phe Ala Leu Ala  
 -107 -105 -100 -95  
 Leu Ile Phe Thr Met Ala Phe Gly Ser Thr Ser Ser Ala Gln Ala  
 -90 -85 -80  
 Ala Gly Lys Ser Asn Gly Glu Lys Lys Tyr Ile Val Gly Phe Lys  
 -75 -70 -65  
 Gln Thr Met Ser Thr Met Ser Ala Ala Lys Lys Lys Asp Val Ile  
 -60 -55 -50  
 Ser Glu Lys Gly Gly Lys Val Gln Lys Gln Phe Lys Tyr Val Asp  
 -45 -40 -35  
 Ala Ala Ser Ala Thr Leu Asn Glu Lys Ala Val Lys Glu Leu Lys  
 -30 -25 -20  
 Lys Asp Pro Ser Val Ala Tyr Val Glu Glu Asp His Val Arg His  
 -15 -10 -5  
 Lys Arg Ala Gln Ser Val Pro Tyr Gly Val Ser Gln Ile Lys Ala  
 1 5 10  
 Pro Ala Leu His Ser Gln Gly Tyr Thr Gly Ser Asn Val Lys Val  
 15 20 25  
 Ala Val Ile Asp Ser Gly Ile Asp Ser Ser His Pro Asp Leu Lys  
 30 35 40  
 Val Ala Gly Gly Ala Ser Met Val Pro Ser Glu Thr Asn Pro Phe  
 45 50 55  
 Gln Asp Asn Asp Ser His Gly Thr His Val Ala Gly Thr Val Ala  
 60 65 70  
 Ala Leu Asn Asn Ser Ile Gly Val Leu Gly Val Ala Pro Ser Ala  
 75 80 85  
 Ser Leu Tyr Ala Val Lys Val Leu Gly Ala Asp Gly Ser Gly Gln  
 90 95 100  
 Asp Ser Trp Ile Ile Asn Gly Ile Glu Trp Ala Ile Ala Asn Asn  
 105 110 115  
 Met Asp Val Ile Asn Met Ser Leu Gly Gly Pro Ser Gly Ser Ala  
 120 125 130  
 Ala Leu Lys Ala Ala Val Asp Lys Ala Val Ala Ser Gly Val Val  
 135 140 145  
 Val Val Ala Ala Ala Gly Asn Glu Gly Thr Ser Gly Ser Ser Ser  
 150 155 160  
 Thr Val Asp Tyr Pro Gly Lys Tyr Pro Ser Val Ile Ala Val Gly  
 165 170 175  
 Ala Val Asp Ser Ser Asn Gln Arg Ala Ser Phe Ser Ser Val Gly  
 180 185 190  
 Pro Glu Leu Asp Val Met Ala Pro Gly Val Ser Ile Gln Ser Thr  
 195 200 205  
 Leu Pro Gly Asn Lys Tyr Gly Ala Tyr Asn Gly Thr Ser Met Ala  
 210 215 220  
 Ser Pro His Val Ala Gly Ala Ala Ala Leu Ile Leu Ser Lys His  
 225 230 235  
 Pro Asn Trp Thr Asn Thr Gln Val Arg Ser Ser Leu Glu Asn Thr  
 240 245 250  
 Thr Thr Lys Leu Gly Asp Ser Phe Tyr Tyr Gly Lys Gly Leu Ile  
 255 260 265  
 Asn Val Gln Ala Ala Ala Gln  
 270 275

## INTERNATIONAL SEARCH REPORT

International Application No  
PL., US 96/02861

A. CLASSIFICATION OF SUBJECT MATTER  
IPC 6 C12N15/57 C12N9/54 C12N1/21 C12P21/06

According to International Patent Classification (IPC) or to both national classification and IPC

## B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)  
IPC 6 C12N C12P

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practical, search terms used)

## C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category *	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	PROTEIN ENGINEERING, vol. 4, no. 7, 1991, pages 719-737, XP002008733 R. SIEZEN ET AL: "Homology modelling and protein engineering strategy of subtilases, the family of subtilisin-like serine proteinases" see the whole document especially figure 3, table V and page 729, right column, last paragraph-page 730 and Discussion ---	1-23, 25-27,29
Y	especially figure 3, table V and page 729, right column, last paragraph-page 730 and Discussion ---	1-29 -/-

Further documents are listed in the continuation of box C.

Patent family members are listed in annex.

## \* Special categories of cited documents :

- "A" document defining the general state of the art which is not considered to be of particular relevance
- "E" earlier document but published on or after the international filing date
- "L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)
- "O" document referring to an oral disclosure, use, exhibition or other means
- "P" document published prior to the international filing date but later than the priority date claimed

"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone

"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art

"Z" document member of the same patent family

1

Date of the actual completion of the international search

18 July 1996

Date of mailing of the international search report

31.07.96

Name and mailing address of the ISA  
European Patent Office, P.B. 5818 Patentlaan 2  
NL - 2280 HV Rijswijk  
Tel. (+ 31-70) 340-2040, Telex 31 651 epo nl.  
Fax (+ 31-70) 340-3016

Authorized officer

Van der Schaaf, C

## INTERNATIONAL SEARCH REPORT

International Application No  
PCT/US 96/02861

## C(Commission) DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Content of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
Y	PROCEEDINGS OF THE NATIONAL ACADEMY OF SCIENCES OF USA, vol. 84, August 1987, WASHINGTON US, pages 5167-5171, XP002008734 J. WELLS ET AL: "Recruitment of substrate-specificity properties from one enzyme into a related one by protein engineering" see the whole document ---	1-29
A	PROCEEDINGS OF THE NATIONAL ACADEMY OF SCIENCES OF USA, vol. 84, March 1987, WASHINGTON US, pages 1219-1223, XP002008735 J. WELLS ET AL: "Designing substrate specificity by protein engineering of electrostatic interactions" see the whole document ---	2-6, 11-13
A	SCIENCE, vol. 233, 1986, pages 659-633, XP002008736 D. ESTELL ET AL: "Probing steric and hydrophobic effects on enzyme-substrate interactions by protein engineering" see the whole document ---	2-6, 11-13
A	ANNALS NEW YORK ACADEMY OF SCIENCES, vol. 672, 1992, pages 71-79, XP002008737 T. GRAYCAR ET AL: "Altering the proteolytic activity of subtilisin through protein engineering" see the whole document ---	2-6
Y	BIOCHEMISTRY 32 (5). 1993. 1199-1203. CODEN: BICHAW ISSN: 0006-2960, XP002008738 RHEINNECKER M ET AL: "ENGINEERING A NOVEL SPECIFICITY IN SUBTILISIN BPN." see the whole document ---	2-6
A	JOURNAL OF BIOLOGICAL CHEMISTRY, vol. 267, no. 23, 15 August 1992, MD US, pages 16335-16340, XP002008739 K. NAKAYAMA ET AL: "Consensus sequence for precursor processing at mono-arginyl sites" see the whole document ---	1,7,24, 25,28,29
1	EP,A,0 316 748 (HOECHST AG) 24 May 1989 see the whole document ---	24,28
3	WO,A,91 11454 (UPJOHN CO) 8 August 1991 see the whole document ---	24,28

-/-

## INTERNATIONAL SEARCH REPORT

International Application No  
PCT/US 96/02861

C(Continued) DOCUMENTS CONSIDERED TO BE RELEVANT		
Category	Citation of document with indication, where appropriate, of the relevant passages	Relevant to claim No.
P,X	WO,A,95 30010 (PROCTER & GAMBLE) 9 November 1995. see claim 1 ---	1-23,26, 27
X	EP,A,0 405 901-(UNILEVER PLC ;UNILEVER NV (NL)) 2 January 1991 see claim 1 ---	1-23,26, 27
P,X	BIOCHEMISTRY 34 (41). 1995. 13312-13319. ISSN: 0006-2960. 17 October 1995, XP002008740 BALLINGER M D ET AL: "Designing subtilisin BPN' to cleave substrates containing dibasic residues." see the whole document ---	7-13, 18-24, 28,29
Y	KEYSTONE SYMPOSIUM ON PROCESSING ENZYMES: BIOCHEMISTRY, GENETICS AND CLINICAL RELEVANCE, LAKE TAHOE, CALIFORNIA, USA, MARCH 2-8, 1995. JOURNAL OF CELLULAR BIOCHEMISTRY SUPPLEMENT 8 (19B). 1995. 233. ISSN: 0733-1959, 4 March 1995, XP002008741 WELLS J A ET AL: "Subtilisin: An enzyme made for engineering." see abstract B7-012 -----	1-6, 14-17, 25-27
P,Y		1-29

**THIS PAGE BLANK (USPTO)**